



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

BAB II

TINJAUAN PUSTAKA

2.1 Pelanggan Churn

Pelanggan adalah individu atau organisasi yang membeli atau menggunakan produk atau layanan dari sebuah perusahaan. Kepentingan pelanggan bagi perusahaan sangat besar karena pelanggan menjadi sumber pendapatan utama. Oleh karena itu, perusahaan harus memahami kebutuhan dan preferensi pelanggan untuk mencapai kesuksesan bisnis. Kata “Churn” berasal dari dua gabungan kata, yaitu “Change” dan “Turn”. Hal ini dapat diartikan bahwa situasi di mana pelanggan ingin mengubah pilihannya terhadap provider dan migrasi dari pelanggan ini harus dihentikan (Yabas dan Cankaya, 2013) (Susan, 2022). Dengan penggabungan dua kata di atas dapat disimpulkan bahwa pelanggan churn adalah seseorang yang sedang berada di posisi yang hendak untuk berpindah ke provider lain dan perpindahan ini harus segera dihentikan untuk menghindari kehilangan pelanggan.

2.2 *Machine learning*

Machine learning adalah salah satu cabang dari kecerdasan buatan (AI) yang menggunakan algoritma untuk mengajarkan komputer untuk belajar dari data dan pengalaman masa lalu, serta membuat prediksi dan keputusan berdasarkan pola dan tren yang teridentifikasi pada data tersebut. *Machine learning* bukan hanya penerapan metode komersial untuk mengekstraksi informasi dari data; belajar juga merupakan syarat kecerdasan. Sistem yang cerdas harus mampu beradaptasi dengan lingkungannya; ia harus belajar untuk tidak mengulangi kesalahannya, tetapi mengulangi keberhasilannya. Agar kecerdasan buatan menjadi kenyataan membutuhkan paradigma baru, jenis pemikiran baru, model komputasi baru, atau rangkaian algoritma baru (Massachusetts Institute of Technology, 2016). Ada tiga jenis *machine learning*, yaitu *unsupervised learning*, *supervised learning*, dan *reinforcement learning*.



2.2.1. *Supervised Learning*

Supervised learning adalah jenis *machine learning* di mana mesin belajar dari *dataset* yang telah diberikan label atau dikategorikan. *Supervised machine learning* melibatkan atribut *output* yang telah ditentukan sebelumnya selain penggunaan atribut *input*. Algoritma ini mencoba untuk memprediksi dan mengklasifikasikan atribut yang telah ditentukan, dan akurasi serta kesalahan klasifikasi mereka di samping ukuran kinerja lainnya, tergantung pada jumlah atribut yang telah ditentukan dan diprediksi dengan benar atau diklasifikasikan atau sebaliknya (Berry *et al.*, 2020). Secara teknis, algoritma yang diawasi melakukan tugas analitik terlebih dahulu menggunakan *training data* dan kemudian membangun fungsi kontingen untuk memetakan *instance* baru dari atribut tersebut. Tujuan dari *supervised learning* adalah untuk membuat model yang dapat memprediksi label atau kategori yang tepat untuk data yang belum pernah dilihat sebelumnya.

2.2.2. *Unsupervised Learning*

Unsupervised learning adalah jenis *machine learning* di mana mesin belajar data *dataset* tanpa label atau kategori. Dalam *unsupervised learning*, mesin diberikan *input* berupa fitur tanpa label dan tujuannya adalah untuk menemukan pola atau struktur yang tersembunyi dalam data. Algoritma *unsupervised learning* cocok untuk membuat label dalam data yang selanjutnya digunakan untuk mengimplementasikan tugas *supervised learning* (Berry *et al.*, 2020). Dengan kata lain, mesin belajar untuk mengelompokkan data ke dalam kelompok atau kluster yang berbeda berdasarkan kesamaan fitur atau kesamaan karakteristik.

Aplikasi *unsupervised learning* yang umum adalah segmentasi pelanggan, di mana mesin belajar untuk mengelompokkan pelanggan berdasarkan kesamaan fitur seperti usia, pendapatan, dan riwayat pembelian tanpa label pelanggan yang spesifik. Hal ini dapat digunakan untuk mengembangkan strategi pemasaran yang lebih efektif. Meskipun tidak memerlukan *dataset* yang terlabel dan dapat menemukan pola tersembunyi dalam data, kelemahan *unsupervised learning* adalah hasilnya hanya berupa kelompok tanpa label, sehingga interpretasi dapat menjadi sulit. Model *unsupervised learning* digunakan untuk *clustering*, *association*, dan *dimensionality reduction*.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



2.2.3. Reinforcement Learning

Reinforcement learning adalah jenis *machine learning* di mana mesin belajar melalui interaksi dengan lingkungannya dan memperoleh pengalaman melalui *trial and error*. Dalam *reinforcement learning*, mesin tidak diberikan label atau kategori seperti pada *unsupervised learning*. Tujuannya adalah untuk membuat model yang dapat mengambil tindakan yang paling tepat untuk mencapai tujuan tertentu. *Reinforcement learning* memiliki beberapa keuntungan, seperti dapat belajar melalui interaksi dengan lingkungan, mengoptimalkan keputusan di lingkungan yang kompleks dan dinamis, serta menghasilkan strategi yang adaptif dan fleksibel. Namun, kelemahan dari *reinforcement learning* adalah membutuhkan waktu dan pengalaman yang cukup untuk belajar dari lingkungan yang kompleks.

2.3 Random Forest Classifier

Random Forest Classifier adalah salah satu metode *ensemble learning* dan algoritma *supervised learning* yang digunakan untuk klasifikasi maupun regresi (Pamina *et al.*, 2019). Seperti namanya, algoritma ini *tree-based* dengan setiap *tree* bergantung pada kumpulan variabel acak (Polikar, 2012). Prediksi churn pelanggan dengan menggunakan pohon keputusan (Decision Trees atau DT) dan regresi logistik (Logistic Regression atau LR) adalah teknik yang sangat populer untuk memperkirakan probabilitas churn, karena keduanya menggabungkan kinerja prediktif yang baik dengan kemampuan pemahaman yang baik (Caigny *et al.*, 2018).

Random Forest Classifier bekerja dengan penggabungan dari banyak pohon keputusan (*decision tree*). Algoritma ini dimulai dengan pemilihan sampel acak, pembuatan pohon keputusan dengan menghitung *split* (pemisahan) yang sesuai berdasarkan nilai-nilai fitur *dataset*, menggabungkan semua hasil pohon dari *decision tree* lalu melakukan “*voting*” atas hasil mayoritas, mereduksi variansi yang mungkin terjadi pada pohon keputusan individu untuk menghindari *overfitting*. Pada proses pembuatan pohon, akan terjadi sebuah percabangan. Percabangan ini ditentukan oleh nilai yang dihasilkan lalu dibandingkan dengan nilai titik pemisah. Jika titik prediktornya lebih kecil daripada nilai titik pemisah, maka variabel tersebut akan memisahkan diri ke cabang kiri. Sementara itu, variabel sisa yang

Hak Cipta :

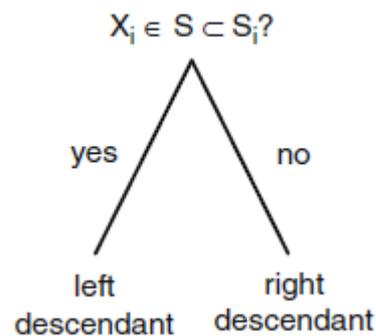
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

bernilai lebih dari nilai titik pemisah, maka akan pergi ke cabang kanan. Lihat Gambar 2.1 untuk visualisasinya (Polikar, 2012).



Gambar 2.1 Visualisasi Percabangan *Decision Tree*

2.4 Metrik Evaluasi

Metrik evaluasi adalah metrik atau angka yang digunakan untuk mengukur efektivitas dan efisiensi model *machine learning* dalam memprediksi atau mengklasifikasikan data. Tujuan dari metrik evaluasi adalah untuk mengukur kemampuan model dalam menghasilkan prediksi yang akurat dan relevan dengan data yang diberikan. Metrik evaluasi ini sangat penting untuk mengevaluasi performa model dan seberapa baik performa model dalam menghadapi data baru yang tidak terlihat.

Beberapa contoh metrik evaluasi umum yang digunakan dalam berbagai tugas *machine learning* adalah Akurasi (*Accuracy*), Presisi (*Precision*), *Recall*, *F1-Score*, dan Area di Bawah Kurva ROC (Area Under the ROC Curve, AUC-ROC). Setiap metrik memiliki interpretasi dan tujuan yang berbeda, dan pemilihan metrik evaluasi yang tepat bergantung pada tugas spesifik dan karakteristik data. Penting untuk mempertimbangkan berbagai metrik evaluasi untuk mendapatkan gambaran lengkap tentang kinerja model dan menghindari kesalahan interpretasi dalam peringkat kinerja model.

2.4.1 *Accuracy Score*

Accuracy Score adalah metrik penilaian yang biasa digunakan dalam tugas klasifikasi pembelajaran mesin. Metrik ini mengukur kemampuan model untuk mengklasifikasikan data dengan benar. Nilai akurasi dihitung dengan cara membagi jumlah prediksi yang benar (koreksi positif dan koreksi negatif) dengan jumlah total



data. Nilai peringkat akurasi berkisar dari 0 hingga 1, di mana 1 berarti model mengklasifikasikan semua data dengan benar dan 0 berarti model tidak memiliki prediksi yang benar sama sekali. Meskipun skor akurasi merupakan metrik yang penting, skor ini harus digunakan dengan hati-hati, terutama dengan kumpulan data yang tidak seimbang, karena tidak akan secara akurat menangkap kelas minoritas dan dapat menyesatkan jika kumpulan data tersebut memiliki ketidakseimbangan kelas yang signifikan. Oleh karena itu, jika terjadi ketidakseimbangan kelas, metrik evaluasi lain seperti skor *precision*, *recall*, atau F1 juga dapat menjadi indikator baik buruknya sebuah model, bergantung pada jenis model yang akan dibuat. Adapun formula untuk menghitung *accuracy score* adalah sebagai berikut (Setyo, 2019):

$$Accuracy = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Total Sampel}} \times 100\% \quad (2.1)$$

2.4.2 F1 Score

Skor F1 adalah metrik evaluasi yang digunakan untuk mengukur kinerja model dalam tugas klasifikasi, terutama ketika ada ketidakseimbangan kelas dalam *dataset*. Metrik ini adalah rata-rata *precision* dan *recall* yang harmonis, sehingga keduanya dapat dihitung secara setara. Skor F1 dihitung menggunakan rata-rata harmonik dari *precision* dan *recall*, dengan skor yang lebih tinggi dihasilkan saat *precision* dan *recall* tinggi. Skor F1 berguna untuk mengukur kemampuan model dalam mengklasifikasikan kelas positif dengan benar sekaligus menghindari kesalahan dalam mengklasifikasikan kelas negatif. Oleh karena itu, skor F1 berguna saat ingin mencapai keseimbangan antara *precision* dan *recall*, terutama di kelas minoritas yang identifikasinya penting. Nilai skor F1 berkisar dari 0 hingga 1, dengan 1 kinerja sempurna dan 0 kinerja buruk. Formula untuk menghitung F1 *Score* sebagai berikut (Resika, 2019):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.2)$$

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



2.4.3 Recall Score

Recall, juga dikenal sebagai *Sensitivity* atau *True Positive Rate*, adalah metrik evaluasi yang digunakan untuk mengukur kemampuan model dalam mengidentifikasi dan mengenali kelas positif yang benar dari semua kelas positif dalam kumpulan data. Metrik ini mengukur jumlah positif aktual yang diidentifikasi model sebagai kelas positif. *Recall* dihitung dengan membagi *True Positive* (TP) dengan jumlah kelas positif (TP + *False Negative*). *Recall* penting dalam tugas-tugas di mana menghindari kesalahan *False Negative* sangat kritis. Skor *recall* berkisar dari 0 hingga 1, dengan 1 mewakili pemulihan sempurna (tidak ada *False Negative*) dan 0 mewakili pemulihan buruk (tidak ada *True Positive*). Formula untuk menghitung *Recall Score* sebagai berikut (Setyo, 2019):

$$\text{Recall} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Negatives}} \quad (2.3)$$

Dimana:

- *True Positives* (TP) adalah jumlah kasus yang benar diprediksi sebagai positif (churn) oleh model
- *False Negatives* (FN) adalah jumlah kasus yang salah diprediksi sebagai negatif oleh model, yang nilai sebenarnya adalah positif (churn)

2.4.4 ROC AUC

ROC AUC (*Receiver Operating Characteristic Area Under the Curve*) adalah metrik evaluasi yang digunakan untuk mengukur kinerja model dalam tugas klasifikasi biner. Kurva ROC adalah grafik yang memplot *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada ambang klasifikasi yang berbeda. Nilai ROC-AUC menunjukkan seberapa baik model dapat membedakan kelas positif dan kelas negatif. ROC-AUC memiliki nilai antara 0 dan 1, dimana nilai 1 menunjukkan hasil model yang sempurna dalam membedakan kedua kelas, dan nilai 0,5 menunjukkan efisiensi yang sama seperti pada klasifikasi acak. Semakin tinggi nilai ROC-AUC, semakin baik model memprediksi kelas positif dan negatif dengan benar. Metrik ini sangat berguna untuk mengukur kinerja model di beberapa ambang klasifikasi, dan sangat relevan untuk kumpulan data yang tidak seimbang di mana bias kesalahan

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, pennisan karya ilmiah, pennisan laporan, pennisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

dapat menjadi kritis. Berikut formula-formula yang digunakan dalam ROC UAC (Datasans, 2019).

FPR dihitung sebagai berikut:

$$FPR = \frac{\text{False Positif}}{\text{True Positif} + \text{True Negatives}} \quad (2.4)$$

TPR (*Recall*) dihitung sebagaimana pada formula (2.3).

$$\text{Recall} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Negatives}} \quad (2.5)$$

2.4.5 Precision

Precision adalah metrik evaluasi yang mengukur sejauh mana model klasifikasi benar dalam mengidentifikasi dan memprediksi data sebagai kelas positif dari semua prediksi yang dibuat sebagai kelas positif. *Precision* mengukur jumlah prediksi positif yang benar (*True Positive*) dibagi dengan jumlah total prediksi positif yang dibuat (*True Positive + False Positive*). *Precision* penting dalam tugas di mana menghindari positif palsu sangat penting. Nilai *Precision* berkisar dari 0 hingga 1, di mana 1 menunjukkan bahwa model membuat prediksi positif yang benar-benar sempurna dan 0 menunjukkan bahwa model sama sekali tidak memiliki prediksi positif yang benar. Semakin tinggi nilai *Precision*, semakin baik model dapat mengidentifikasi data dengan benar sebagai kelas positif. Formula untuk menghitung *Precision* sebagai berikut (Setyo, 2019):

$$\text{Precision} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Positive}} \quad (2.6)$$

2.5 Churn Prediction

Prediksi churn pelanggan adalah bagian dari manajemen churn, yang memprediksi perilaku pelanggan dengan klasifikasi sebagai pelanggan setia dan mana yang cenderung untuk pindah ke kompetitor lain. “Pelanggan churn” berarti



© Hak Cipta milik Politeknik Negeri Jakarta

kehilangan klien (Zeniarta dan Luthfiarta, 2015). Tujuan dari *churn prediction* adalah untuk membantu perusahaan dalam menjaga kepuasan pelanggan dan meminimalkan kehilangan pelanggan. Dalam proses *churn prediction*, data pelanggan seperti data penggunaan layanan, data pembayaran, dan data sosial media digunakan untuk membangun model prediksi churn menggunakan teknik *machine learning*, seperti *logistic regression*, *decision tree*, dan *neural network*. *Churn prediction* termasuk ke dalam kategori *supervised learning*. Hal ini dikarenakan dalam *churn prediction* memerlukan data training yang sudah dilabeli atau ditandai dengan label churn. Penerapan *churn prediction* pada perusahaan telah terbukti efektif dalam membantu perusahaan dalam meminimalkan churn dan meningkatkan retensi pelanggan. Dalam sebuah studi yang dilakukan oleh McKinsey, ditemukan bahwa perusahaan yang mengimplementasikan *churn prediction* mampu meningkatkan retensi pelanggan hingga 10-15% dan meningkatkan profitabilitas perusahaan hingga 20-25%

2.6 Website

2.6.1 Front-End

Front-End merupakan salah satu bagian terpenting dalam berdirinya sebuah *website*. *Front-End* adalah bagian yang langsung dapat terlihat oleh *user*, sisi tampilan. Bagian ini dibuat dengan menggunakan HyperText Markup Language (HTML), Cascading Style Sheets (CSS), dan juga JavaScript. Dengan komponen – komponen tersebut, gambar, teks, warna menu, tombol, dan interaksi lainnya terbuat dan dapat dilihat atau digunakan oleh *user*.

2.6.2 Back-End

Back-End adalah bagian dari aplikasi atau *website* yang tidak terlihat oleh pengguna akhir. *Back-End* terdiri dari server, database, dan aplikasi server yang mengelola logika bisnis, pemrosesan data, dan operasi server. Pada umumnya Back-End development melibatkan pemrograman dengan bahasa pemrograman Python, Ruby, PHP, atau Java, serta menggunakan database seperti MySQL, PostgreSQL, atau MongoDB. Peran *Back-End development* semakin penting dengan semakin kompleksnya aplikasi dan *website* yang dibangun. *Back-End developer* bertanggung jawab untuk memastikan bahwa server dan database dapat

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

menangani beban trafik yang tinggi dan menjaga konsistensi dan keamanan data. *Back-End development* bekerja sama dengan *Front-End development* untuk menciptakan aplikasi atau *website* yang optimal dari segi fungsionalitas dan tampilan

2.7 Python

Python adalah bahasa pemrograman tingkat tinggi yang sangat terkenal dan digunakan secara luas di berbagai bidang. Bahasa ini ditandai dengan sintaksis yang simpel, mudah dibaca, dan efisien, sehingga mudah dipelajari oleh pemula sekalipun. Python memiliki pendekatan yang kuat terhadap kesederhanaan dan ekspresivitas, memungkinkan pengembang untuk menulis kode yang efisien dan singkat. Selain itu, Python memiliki banyak *library* dan modul yang luas dan beragam, yang memperluas kemampuan bahasa ini hingga mencakup berbagai domain seperti ilmu data, *machine learning*, pengembangan web, analisis data, dan banyak lagi.

Salah satu keunggulan besar Python adalah komunitas pengguna yang besar dan aktif. Python memiliki komunitas yang luas dari pengembang, mahasiswa, ilmuwan data, dan profesional di berbagai industri. Komunitas ini memberikan dukungan, sumber daya, dan kontribusi yang melimpah, menjadikan Python sebagai lingkungan yang ideal untuk belajar dan bekerja sama. Komunitas Python juga terkenal dengan sikap inklusif dan ramah, menjadikannya lingkungan yang mendukung bagi siapa pun yang ingin belajar dan berkembang dalam dunia pemrograman.

Di samping itu, Python juga didukung oleh proyek *open-source* yang tangguh dan *library* pihak ketiga yang melimpah. *Library* seperti NumPy, Pandas, Matplotlib, TensorFlow, dan Flask telah menjadi dasar dari berbagai proyek dan aplikasi, mulai dari analisis data hingga pembuatan aplikasi web. Python juga merupakan bahasa yang banyak digunakan di lingkungan akademik dan riset karena fleksibilitasnya dalam menangani masalah yang rumit dan kemampuannya untuk mengintegrasikan teknologi canggih seperti *machine learning* dan kecerdasan buatan. Dengan kombinasi kesederhanaan, fleksibilitas, dukungan komunitas yang tangguh, dan ekosistem *library* yang kaya, Python menjadi salah satu bahasa



pemrograman yang paling diminati dan berpengaruh dalam industri teknologi dan ilmu pengetahuan.

2.8 *Library*

Library Python, atau lebih dikenal sebagai modul, adalah kumpulan kode yang telah dibuat dan diatur dalam *file* terpisah untuk menyediakan fungsi-fungsi dan alat-alat yang siap pakai bagi pengembang Python (algoritma, 2022). Python memiliki ekosistem *library* yang sangat kaya dan luas, yang mencakup ribuan modul yang mencakup berbagai fitur dan fungsionalitas. *Library-library* ini memperluas kemampuan Python hingga mencakup berbagai domain seperti pemrosesan data, analisis statistik, kecerdasan buatan, pemrograman web, pemrosesan gambar, dan banyak lagi. Beberapa *library* Python yang terkenal dan banyak digunakan adalah NumPy, pandas, Matplotlib, TensorFlow, Flask, dan Django.

Library Python memainkan peran penting dalam mempercepat proses pengembangan perangkat lunak dan mengurangi beban kerja pengembang dalam membuat fungsi-fungsi dari awal. Dengan menggunakan *library*, pengembang dapat dengan mudah mengimpor dan menggunakan fungsi-fungsi yang sudah ada, sehingga tidak perlu menulis kode dari awal untuk tugas-tugas yang umum dan terstandarisasi. Selain itu, *library* Python juga memungkinkan pengembang untuk menghindari kesalahan umum dan memastikan kualitas kode, karena *library-library* tersebut telah diuji secara luas dan diandalkan oleh komunitas Python yang besar.

Komunitas Python yang aktif juga berperan dalam perkembangan dan pemeliharaan *library-library* ini. Pengembang dari seluruh dunia berkontribusi dalam perkembangan *library-library* tersebut, menambahkan fitur-fitur baru, memperbaiki kesalahan, dan memastikan *library* tetap relevan dengan perkembangan teknologi. Oleh karena itu, *library* Python terus berkembang dan mengikuti tren terbaru dalam dunia teknologi, menjadikannya sebagai salah satu kekuatan utama dalam popularitas dan daya tarik bahasa pemrograman Python. Adapun *library-library* yang akan digunakan pada skripsi ini akan dibahas setelah ini.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



2.8.1 Streamlit

Streamlit adalah sebuah framework *open-source* yang diperuntukkan Machine Learning dan Data Science (Streamlit, 2019). Streamlit adalah *library* Python yang memudahkan pembuatan aplikasi web interaktif untuk analisis dan visualisasi data. Dengan Streamlit, pengembang dapat dengan cepat mengubah kode Python menjadi aplikasi web yang menampilkan data dari hasil analisis dalam bentuk bagan, tabel, dan elemen interaktif lainnya. Streamlit menawarkan antarmuka yang sederhana dan ramah, sehingga pemula pun dapat dengan cepat membuat aplikasi web yang menarik dan bermanfaat.

2.8.2 Pandas

Pandas adalah *library* yang cepat, *powerful*, fleksibel, dan mudah digunakan untuk analisis data dan alat manipulasi (Pandas, 2009). Pandas adalah *library* Python yang populer untuk analisis data. *Library* ini menyediakan struktur data dan alat analisis yang kuat untuk membantu pengguna memproses dan menganalisis data. Panda mendukung struktur data seperti DataFrame dan Series, yang memungkinkan pengguna dengan mudah melakukan operasi seperti memfilter, mengelompokkan, menggabungkan, menggabungkan, dan mengubah data. Panda juga terintegrasi dengan baik dengan banyak *library* analisis data lainnya, menjadikannya pilihan terbaik bagi pengguna yang melakukan tugas analisis data dengan Python.

2.8.3 LazyPredict

Library LazyPredict adalah *library* Python yang menyediakan alat otomatisasi untuk menguji dan mengevaluasi berbagai model *machine learning*. *LazyPredict* memungkinkan pengguna untuk dengan mudah mengevaluasi berbagai model *machine learning* tanpa harus menulis kode yang rumit dan membosankan. *Library* ini secara otomatis melatih sejumlah besar model *machine learning* yang umum digunakan dengan pengaturan dan *hyperparameter* yang berbeda, serta menghasilkan ringkasan performa untuk setiap model (Lewinson, 2021). Hasil diberi peringkat berdasarkan kinerja terbaik, memungkinkan pengguna dengan mudah memilih model yang paling cocok untuk tugas tertentu.

LazyPredict sangat berguna untuk pemula yang ingin membandingkan berbagai model dan melihat bagaimana mereka melakukan tugas tertentu. Dengan

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

sejumlah kecil kode, pengguna dapat dengan cepat mendapatkan gambaran umum tentang kinerja berbagai model, yang dapat membantu mereka membuat keputusan tentang model yang paling sesuai untuk proyek *machine learning* mereka. Dengan mempercepat evaluasi model, *LazyPredict* juga dapat meningkatkan efisiensi dan produktivitas pembuatan model *machine learning*.

2.8.4 Matplotlib

Matplotlib merupakan suatu *library* yang komprehensif untuk membuat visualisasi statis, animasi, dan interaksi (Matplotlib, 2003). *Library* ini menawarkan berbagai fungsi dan metode untuk membuat grafik dan visualisasi data dalam format yang informatif dan menarik. Matplotlib mendukung berbagai jenis bagan, seperti *scatter plot*, *line plot*, *bar plot*, *histogram*, dll. Matplotlib memungkinkan pengguna untuk dengan mudah menyesuaikan tampilan bagan dengan mengubah warna, label, dan gaya visual lainnya. Matplotlib juga dapat diintegrasikan dengan *library* analisis data lainnya seperti Pandas, sehingga memudahkan pengguna untuk membuat visualisasi data yang informatif.

2.8.5 Pickle

Pickle adalah *library* Python yang digunakan untuk membuat serial dan membuat serial objek Python ke dalam format yang dapat disimpan ke *file* atau dikirim melalui web (GeeksforGeeks, 2018). Pickle memungkinkan pengguna untuk menyimpan objek Python ke *file* dan kemudian mengambilnya dari *file* dalam format penyimpanannya. Pickle sangat berguna untuk menyimpan dan memuat ulang objek kompleks, seperti model *machine learning* yang telah dilatih atau data yang membutuhkan waktu lama untuk diproses. Selain itu, Pickle juga dapat digunakan untuk melewati objek Python di antara berbagai proses atau mesin, menjadikannya pilihan populer di lingkungan komputasi terdistribusi.

2.8.6 Seaborn

Seaborn adalah *library* visualisasi data Python yang berdasarkan dari *library* Matplotlib (Seaborn, 2012). *Library* ini didasarkan pada Matplotlib dan menyediakan fungsi lanjutan yang mudah digunakan untuk membuat visualisasi yang elegan. Seaborn menawarkan fitur-fitur seperti bagan berbasis model, *heatmap*, *pair plot*, dan lainnya. Seaborn memungkinkan pengguna untuk dengan



mudah membuat visualisasi statistik yang menarik dan informatif, bahkan dengan kode yang lebih sedikit dibandingkan saat menggunakan Matplotlib secara langsung.

2.8.7 Six

Library Six adalah *library* Python yang memfasilitasi migrasi kode dari Python 2 ke Python 3. Saat berpindah dari Python 2 ke Python 3, ada perbedaan signifikan dalam sintaks dan perilaku bahasa, mengakibatkan kode Python 2 tidak selalu berfungsi dengan benar di Python 3. *Library Six* menawarkan beberapa fungsi kompatibilitas yang memungkinkan penulisan dan dekorasi kode. kompleksitas proses transfer. *Six* memungkinkan pengguna untuk dengan mudah menulis kode Python yang berfungsi di kedua versi juru bahasa, meningkatkan portabilitas dan mendukung fitur Python 3 terbaru tanpa merusak kompatibilitas Python 2. *Library Six* adalah alat berguna yang membantu pengembang dan organisasi berpindah dari Python 2 ke Python 3 dengan lebih mulus dan efisien. NumPy adalah *library* Python yang digunakan untuk perhitungan numerik dan operasi array multidimensi. *Library* ini menyediakan struktur data array yang kuat dan fungsi matematika yang kuat yang memungkinkan pengguna melakukan operasi numerik dengan cepat dan mudah. NumPy menjadi dasar bagi banyak *library* lain di ekosistem Python, seperti Pandas, Matplotlib, dan Scikit-Learn. Dengan NumPy, pengguna dapat dengan mudah melakukan perhitungan numerik, manipulasi tabel, dan manipulasi data dalam format vektor atau matriks.

2.8.8 Joblib

Joblib adalah *library* Python yang memungkinkan pemrosesan paralel dan caching kode Python. Joblib menyediakan pipeline yang ringan dan berjalan pada bahasa pemrograman Python (Joblib, 2009). *Library* ini sering digunakan untuk mempercepat pemrosesan yang membutuhkan perhitungan besar atau berulang. Joblib memungkinkan pengguna untuk dengan mudah memproses hasil kalkulasi secara paralel dan menyimpannya untuk digunakan nanti, menghemat waktu dan sumber daya pada kode yang memerlukan kalkulasi berulang.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



2.8.9 Sys

Sys adalah modul *library* standar Python yang menyediakan akses ke sekumpulan variabel dan fungsi yang berinteraksi langsung dengan interpreter Python dan sistem operasi. Modul Sys memungkinkan pengguna untuk mengakses argumen baris perintah, mengatur dan mengambil informasi tentang lingkungan Python, dan mengontrol *input* dan *output* stream. Modul Sys juga memungkinkan penggunaan variabel dan fungsi sistem, membuatnya berguna untuk mengelola program dan lingkungan eksekusi.

2.8.10 Os

Os adalah modul Python yang menyediakan berbagai fungsi untuk berinteraksi dengan sistem operasi saat ini. Modul ini memungkinkan pengguna untuk membuat, menghapus, dan mengedit folder dan *file*, mengatur izin *file*, menjalankan *command shell*, dan banyak lagi. Modul Os memungkinkan pengguna untuk melakukan tugas manajemen langsung dari kode Python, membuat aplikasi yang terintegrasi lebih baik dengan sistem operasi.

2.8.11 Io

Modul Io dari Python Standard *Library* menyediakan fungsi untuk mengakses dan memanipulasi *input* dan *output* (I/O) stream. Modul ini menyediakan antarmuka terpadu untuk berinteraksi dengan berbagai jenis aliran masukan dan keluaran, seperti *file* teks, aliran *byte* atau objek seperti string karakter dan *byte*. Dengan menggunakan modul Io, pengguna dapat dengan mudah melakukan operasi I/O dengan Python tanpa mengkhawatirkan jenis aliran atau platform yang digunakan.

2.8.12 Imbalanced-Learn

Imbalanced-Learn adalah suatu *open-source library* yang berlisensi MIT yang mengandalkan Scikit-Learn dan menyediakan alat untuk menangani klasifikasi dengan kelas yang tidak seimbang (imbalanced-learn, 2014). Masalah ketidakseimbangan kelas terjadi ketika jumlah contoh di satu kelas jauh lebih sedikit daripada jumlah contoh di kelas lain. Masalah ini sering terjadi di banyak tugas klasifikasi dunia nyata, seperti deteksi penipuan, deteksi penyakit langka, dan

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



deteksi anomali, di mana kelas minoritas sering kali lebih besar daripada kelas mayoritas.

2.8.13 NumPy

NumPy adalah *library* Python yang digunakan untuk perhitungan numerik dan operasi *array* multidimensi. *Library* ini menyediakan struktur data *array* yang kuat dan fungsi matematika yang kuat yang memungkinkan pengguna melakukan operasi numerik dengan cepat dan mudah. NumPy menghadirkan kekuatan komputasi bahasa, seperti C, Fortran, hingga Python yang lebih mudah dipelajari dan digunakan (NumPy, 2006). NumPy menjadi dasar bagi banyak *library* lain di ekosistem Python, seperti Pandas, Matplotlib, dan Scikit-Learn. Dengan NumPy, pengguna dapat dengan mudah melakukan perhitungan numerik, manipulasi tabel, dan manipulasi data dalam format vektor atau matriks.

2.8.14 PyPDF2

PyPDF2 adalah *library* Python untuk membaca, mengedit, dan membuat *file* PDF. *Library* ini menawarkan berbagai fungsi untuk menggabungkan, memisahkan, dan memutar halaman PDF, serta mengekstraksi teks dan gambar dari *file* PDF (PyPDF2, 2011). PyPDF2 memungkinkan pengguna untuk dengan mudah melakukan operasi terkait PDF dengan Python, menjadikannya pilihan yang berguna untuk proyek yang berhubungan dengan *file* PDF.

2.8.15 Google Cloud Storage

Google Cloud Storage adalah *library* Python yang memungkinkan interaksi dengan penyimpanan berbasis *cloud* milik Google, yaitu Google Cloud Storage. *Library* ini memungkinkan pengguna mengunggah dan mengunduh *file*, membuat dan menghapus *Bucket* (wadah penyimpanan), dan melakukan operasi lain di layanan penyimpanan Google Cloud Storage. *Library* ini memfasilitasi integrasi aplikasi Python ke dalam infrastruktur *cloud* Google, yang dapat digunakan untuk menyimpan dan mengelola data di lingkungan *cloud*.

2.8.16 File Connection

File Connection adalah *library* Python yang memungkinkan pengguna untuk membuka, membaca, menulis, dan mengelola *file* di sistem operasi. *Library* ini menyediakan fungsi untuk membuka dan menutup *file*, serta membaca dan

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

menulis data dalam *file*. *Library File Connection* memungkinkan pengguna untuk dengan mudah melakukan operasi I/O pada *file*, memungkinkan data disimpan dan diambil dari *file* di lingkungan Python.

2.9 GitHub

GitHub merupakan manajemen proyek, sistem *versioning code*, sekaligus platform jaringan sosial bagi para developer seluruh dunia (Dicoding, 2021). GitHub adalah platform *hosting* dan kolaborasi yang sangat populer untuk perangkat lunak berbasis Git. Git adalah sistem kontrol versi terdistribusi yang digunakan untuk melacak perubahan kode sumber selama pengembangan perangkat lunak. GitHub memungkinkan pengembang untuk dengan mudah menyimpan dan membagikan kode sumber mereka, mengelola perubahan antar cabang, dan berkolaborasi dalam tim untuk mengembangkan proyek. Selain itu, platform ini menawarkan alat kolaborasi yang efektif seperti masalah dan permintaan penarikan yang memungkinkan diskusi terstruktur dan terorganisasi serta penyelesaian masalah perangkat lunak.

GitHub adalah *hub* untuk banyak proyek sumber terbuka dan perangkat lunak sumber terbuka lainnya di seluruh dunia. Ini memfasilitasi partisipasi dan masukan dari komunitas pengembang besar yang dapat membantu mengidentifikasi *bug*, menyarankan perbaikan, atau menambahkan fitur baru ke proyek yang sudah ada. GitHub juga memainkan peran penting dalam memfasilitasi kolaborasi antara pengembang, tim, dan organisasi untuk mengelola proyek kode dan perangkat lunak, menjadikan platform ini salah satu tempat paling populer dan penting bagi pengembang.

2.10 Hyperparameter Tuning

Hyperparameter tuning adalah tahap untuk menemukan kombinasi terbaik dari parameter tambahan dalam model *machine learning* agar meningkatkan kinerja dan akurasi model. Hyperparameter tuning adalah proses mengidentifikasi tuple hyperparameters yang menghasilkan model optimal yang meminimalkan fungsi kerugian yang telah ditentukan pada data independen yang disediakan (CodingStudio, 2023). Parameter tambahan ini tidak diajarkan oleh model, tetapi harus ditentukan oleh pengguna atau pembuat model sebelum melatih model.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Contoh parameter tambahan termasuk tingkat pembelajaran (*learning rate*), jumlah lapisan dan unit dalam jaringan saraf (*neural network*), nilai k , dan lain-lain. Adapun teknik-teknik yang sering digunakan pada algoritma klasifikasi akan dibahas sebagai berikut.

2.10.1 Grid Search

Grid search adalah salah satu teknik untuk mencari kombinasi *hyperparameter* yang optimal untuk model *machine learning*. Grid search digunakan untuk mencari *hyperparameter* yang optimal untuk sebuah model (Joseph, 2018). Metode ini melibatkan penelusuran secara sistematis melalui semua kombinasi nilai yang telah ditentukan sebelumnya untuk setiap *hyperparameter* yang ingin dioptimalkan. Proses *Grid Search* berfungsi sebagai eksplorasi seluruh ruang parameter yang mungkin, dengan mencoba semua kombinasi nilai yang berbeda untuk setiap *hyperparameter*, dan memilih kombinasi yang memberikan kinerja terbaik berdasarkan metrik evaluasi yang ditentukan, seperti tingkat ketepatan, keakuratan, atau skor F1.

2.10.2 Random Search

Random Search adalah salah satu teknik penyetelan *hyperparameter* yang digunakan untuk mencari kombinasi nilai *hyperparameter* yang terbaik dalam model *machine learning*. Random search suatu metode di mana kombinasi acak dari *hyperparameter* dipilih dan digunakan untuk melatih model (Lalithnarayan, 2021). Berbeda dengan *Grid Search* yang melakukan pemindaian sistematis pada semua kombinasi nilai *hyperparameter*, *random search* secara acak memilih kombinasi nilai untuk dievaluasi. Pendekatan ini memiliki keuntungan dalam efisiensi waktu dan sumber daya, terutama ketika terdapat banyak nilai *hyperparameter* dan tidak semua kombinasi nilai memiliki dampak signifikan pada kinerja model. Dalam *Random Search*, dapat menentukan rentang nilai *hyperparameter* dan jumlah iterasi yang ingin dilakukan. Pada setiap iterasi, nilai *hyperparameter* dipilih secara acak dari rentang yang ditentukan. Setelah melakukan evaluasi menggunakan metrik kinerja yang ditentukan (seperti keakuratan atau skor F1) pada data validasi, dapat mencatat kombinasi nilai yang memberikan kinerja terbaik. Proses ini diulang sejumlah iterasi yang telah ditentukan sebelumnya, dan pada akhirnya memilih kombinasi nilai yang



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

memberikan kinerja terbaik dari semua iterasi sebagai kombinasi *hyperparameter* yang optimal untuk model.

Random Search sering lebih efisien daripada *Grid Search*, terutama ketika ada banyak *hyperparameter* dan iterasi terbatas. Pendekatan acak ini memungkinkan variasi dan eksplorasi yang lebih banyak dalam *hyperparameter*, sehingga lebih mungkin menemukan kombinasi nilai yang lebih baik tanpa harus mencoba semua kemungkinan seperti pada *Grid Search*. Namun, tergantung pada masalah dan data yang spesifik, ada kemungkinan bahwa Cari acak tidak dapat menemukan kombinasi nilai yang optimal seperti yang bisa dilakukan oleh *Grid Search*. Oleh karena itu, pemilihan metode *hyperparameter tuning* yang tepat harus mempertimbangkan *trade-off* antara eksplorasi acak dan eksplorasi yang lebih sistematis seperti yang dilakukan oleh *Grid Search*.

2.10.3 Bayesian Optimization

Bayesian Optimization adalah teknik *hyperparameter tuning* yang menggunakan pendekatan probabilistik untuk mencari kombinasi nilai *hyperparameter* yang optimal. Teknik ini memberikan prinsip berdasarkan Bayes Theorem untuk mengarahkan pencarian dari masalah *global optimization* yang efisien dan efektif (Brownlee, 2020). Pendekatan ini menggabungkan pemahaman tentang kinerja model yang telah diuji dengan pendekatan cerdas untuk mencari kombinasi terbaik berdasarkan hasil sebelumnya. Metode ini bekerja berdasarkan model probabilitas yang menggambarkan hubungan antara *hyperparameter* dan kinerja model. Dengan menggunakan model ini, *Bayesian Optimization* dapat menentukan *hyperparameter* mana yang paling mungkin memberikan kinerja terbaik, sehingga lebih efisien daripada metode pencarian acak seperti *random search*.

Dalam *Bayesian Optimization*, memulai dengan memilih beberapa nilai *hyperparameter* awal secara acak dan melatih model menggunakan nilai-nilai tersebut. Setelah melatih model, menggunakan hasil evaluasi untuk membangun model probabilistik yang memodelkan hubungan antara *hyperparameter* dan kinerja model. Model ini disebut model pengganti, dan model ini akan digunakan untuk memprediksi kinerja model untuk kombinasi *hyperparameter* yang belum



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

diuji. Selanjutnya, menggunakan teknik heuristik seperti peningkatan yang diharapkan atau probabilitas peningkatan untuk menentukan kombinasi *hyperparameter* yang paling mungkin memberikan peningkatan kinerja terbaik. Setelah itu, menguji kembali kombinasi nilai *hyperparameter* yang baru, dan proses ini diulang beberapa kali hingga dapat menemukan kombinasi nilai yang memberikan kinerja terbaik. Dengan cara ini, *Bayesian Optimization* secara cerdas memilih kombinasi nilai *hyperparameter* yang paling potensial untuk mencapai kinerja terbaik, sehingga mempercepat proses pencarian penyetelan *hyperparameter* dengan lebih efisien.

2.10.4 Evolutionary Algorithms

Evolutionary Algorithms adalah teknik *hyperparameter tuning* yang terinspirasi oleh teori evolusi dalam biologi. Teknik ini adalah pendekatan berbasis heuristik untuk menyelesaikan masalah yang tidak dapat diselesaikan dengan mudah dalam waktu polinomial (Soni, 2018). Pendekatan ini menggunakan konsep perubahan dan seleksi alam untuk mencari kombinasi nilai *hyperparameter* yang optimal dalam model *machine learning*. Proses ini mirip dengan cara populasi makhluk hidup beradaptasi dan berkembang di alam untuk menghadapi perubahan lingkungan. Dalam *Evolutionary Algorithms*, dimulai dengan menghasilkan populasi awal dari beberapa solusi potensial yang acak. Setiap solusi dalam populasi mewakili satu kombinasi nilai *hyperparameter*. Selanjutnya, mengevaluasi performa model untuk setiap solusi dalam populasi menggunakan metrik evaluasi yang telah ditentukan. Solusi-solusi yang memberikan performa terbaik akan dipilih untuk digunakan dalam proses evolusi selanjutnya.

Dalam proses evolusi, menerapkan operasi genetik seperti reproduksi, mutasi, dan persilangan pada solusi-solusi yang dipilih. Reproduksi adalah proses menghasilkan salinan dari solusi yang unggul, sedangkan mutasi adalah proses mengubah solusi secara acak untuk memperkenalkan variasi. Persilangan adalah proses menggabungkan dua solusi yang berbeda untuk menghasilkan solusi baru yang mungkin lebih baik. Dengan menerapkan operasi genetik ini, menciptakan generasi baru dari solusi yang kemudian akan diuji kembali dan dievaluasi untuk memilih solusi terbaik. Proses evolusi ini diulang sejumlah iterasi yang telah ditentukan sebelumnya atau hingga mencapai kriteria konvergensi. Dengan cara ini,



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritikan atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Evolutionary Algorithms mencari kombinasi nilai *hyperparameter* yang paling cocok secara adaptif dan dengan pendekatan yang mirip dengan proses evolusi alam, sehingga memungkinkan menemukan kombinasi nilai yang lebih baik tanpa harus mencoba semua kemungkinan kombinasi seperti pada *Grid Search*.

2.11 Cloud Computing

Cloud Computing adalah sebuah model yang memungkinkan akses jaringan yang merata, nyaman, dan *on-demand* ke sekumpulan sumber daya komputasi yang dapat dikonfigurasi secara bersama (seperti jaringan, server, penyimpanan, aplikasi, dan layanan) yang dapat dengan cepat diberikan dan dilepaskan dengan usaha manajemen yang minimal atau interaksi dengan penyedia layanan (Mell dan Grance, 2011). Dengan layanan *cloud*, sumber daya seperti server, penyimpanan, basis data, jaringan, dan perangkat lunak disediakan sebagai layanan yang dapat diakses pengguna saat dibutuhkan. Model ini memungkinkan pengguna untuk menyewa dan menggunakan sumber daya komputasi sesuai kebutuhan tanpa memiliki infrastruktur fisik atau mengkhawatirkan teknis penggunaan dan pengelolaan sumber daya.

Keuntungan utama layanan *cloud* adalah fleksibilitas dan skalabilitasnya yang tinggi, yang memungkinkan organisasi dan pengguna mengalokasikan dan menyebarkan sumber daya sesuai dengan kebutuhan bisnis mereka. Selain itu, komputasi *cloud* juga mengurangi biaya investasi awal dalam infrastruktur fisik dan operasional, karena pengguna hanya membayar sumber daya yang benar-benar digunakan. Layanan *cloud* juga sering menawarkan berbagai tingkatan harga dan model penagihan berbasis penggunaan, yang memungkinkan pengguna untuk memilih dan membayar hanya untuk sumber daya yang benar-benar mereka gunakan, menjadikannya lebih hemat biaya. Dengan menggunakan *Cloud Computing*, perusahaan dapat dengan mudah *scaling* dan mengurangi kapasitas sesuai dengan kebutuhan bisnis, mempercepat waktu implementasi proyek, dan memanfaatkan teknologi canggih tanpa harus berurusan dengan infrastruktur TI yang rumit secara internal.



2.12 Google Cloud Storage

Google Cloud Storage adalah layanan penyimpanan berbasis *cloud* yang disediakan oleh Google Cloud Platform (GCP). Layanan ini memungkinkan pengguna untuk menyimpan dan mengelola data dalam jumlah besar secara aman dan efisien di *cloud*. Google Cloud Storage menawarkan fleksibilitas dan skalabilitas yang tinggi, sehingga cocok untuk berbagai kebutuhan penyimpanan, mulai dari arsip, database, konten media, hingga data sensor IoT. Data yang disimpan di Google Cloud Storage dapat diakses dari mana saja dengan koneksi Internet, dan GCP menawarkan berbagai opsi keamanan untuk melindungi data dari akses atau kehilangan yang tidak sah.

Google Cloud Storage juga menawarkan beberapa API dan alat manajemen data yang canggih, seperti: penyimpanan objek, penyimpanan arsip, dan penyimpanan data berkecepatan tinggi. Layanan ini memiliki tingkatan harga yang berbeda dan opsi penagihan berdasarkan penggunaan sehingga pengguna dapat memilih opsi yang sesuai dengan kebutuhan mereka. Selain itu, Google Cloud Storage juga terintegrasi dengan beberapa layanan lain yang disediakan oleh Google Cloud Platform seperti BigQuery, Google Cloud Dataflow dan lainnya, yang memungkinkan pengguna untuk melakukan analisis data yang canggih dan terintegrasi menggunakan data yang tersimpan di Google Cloud Storage.

2.13 Pre-processing Data

Pre-processing data merupakan langkah penting dalam analisis data dan membangun model *machine learning*. Tujuan utamanya adalah untuk membersihkan data mentah, menyiapkannya, dan mengubahnya menjadi format yang lebih sesuai dan relevan untuk analisis lebih lanjut. Tujuan dari langkah ini adalah untuk meningkatkan kualitas data, menghilangkan kesalahan, menghilangkan ketidaklengkapan atau *outlier*, dan memperbaiki potensi masalah ketidakseimbangan kelas.

Pre-processing data biasanya mencakup langkah-langkah seperti menghapus nilai yang hilang atau mengisi nilai yang hilang, menormalkan atau menstandarkan data untuk menghaluskan skala, mengidentifikasi dan menghapus *outlier*, menyelesaikan masalah ketidakseimbangan kelas melalui *oversampling*

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



atau *undersampling*, dan mengubah data menjadi format yang sesuai untuk model *input*. Pra-pemrosesan yang baik memastikan bahwa data siap untuk digunakan dengan benar, berkontribusi secara signifikan terhadap kualitas analisis data dan akurasi model, serta membantu menghindari bias atau kesalahan yang mungkin terjadi selama fase analisis atau konstruksi model *machine learning*. Proses yang terdapat pada *pre-processing data* antara lain, *data integration*, *data transformation*, *Feature Engineering*, *data reduction*, dan *data normalization*. Proses-proses tersebut akan dijelaskan sebagai berikut.

2.13.1 Data Integration

Data integration adalah proses menemukan, memindahkan, dan menggabungkan data dari berbagai sumber untuk memperkaya wawasan dan memperkuat *machine learning* (Trias, 2020) (Lenzerini dan Maurizio, 2002). Integrasi data atau *data integration* ialah proses menggabungkan data dari berbagai sumber yang berbeda menjadi satu kesatuan yang seragam dan terpadu. Tujuan utama dari integrasi data adalah untuk menciptakan kumpulan data yang komprehensif dan terperinci, sehingga memungkinkan pengguna atau sistem untuk mengakses, menganalisis, dan memanfaatkan informasi dari berbagai sumber dengan lebih efisien dan efektif. Proses ini melibatkan transformasi data untuk menyamakan format dan struktur data dari sumber yang berbeda, serta menghilangkan duplikasi data agar kualitas dan konsistensi informasi tetap terjaga. integrasi data dapat dilakukan secara manual oleh manusia, tetapi sering menggunakan alat dan teknologi otomatis, seperti *Extract, Transform, Load* (ETL) atau platform integrasi data, untuk *automize* proses penggabungan dan transformasi data.

Integrasi data menjadi semakin penting seiring dengan meningkatnya kompleksitas dan volume data yang dihasilkan oleh berbagai sumber, seperti aplikasi bisnis, sensor IoT, platform media sosial, dan banyak lagi. Dengan mengintegrasikan data dari berbagai sumber, suatu organisasi dapat mendapatkan wawasan baru yang tidak mungkin ditemukan sebelumnya. Selain itu, integrasi data juga membantu mengurangi kumpulan data yang dapat menyebabkan duplikasi dan kesulitan dalam mengakses informasi yang penting. Dengan data yang terintegrasi dengan baik, suatu organisasi dapat mengoptimalkan proses pengambilan

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



keputusan, meningkatkan analisis data, dan meningkatkan pemahaman tentang pelanggan dan operasi bisnis.

Namun, integrasi data juga menantang karena perbedaan dalam format, struktur, dan makna data dari berbagai sumber. Selain itu, skalabilitas dan kecepatan dalam mengintegrasikan data juga menjadi pertimbangan penting, terutama ketika berurusan dengan data yang sangat besar dan berubah dengan cepat. Oleh karena itu, pemilihan alat dan strategi integrasi data yang tepat sangat penting untuk kesuksesan proses ini. Dengan solusi yang baik, organisasi dapat mencapai pengambilan keputusan yang lebih baik, penerapan strategi yang lebih efektif, dan respons yang lebih cepat terhadap perubahan pasar, yang semuanya memberikan keuntungan kompetitif yang berarti di era informasi yang terus berkembang.

2.13.2 Data Transformation

Transformasi data atau *data transformation* adalah proses mengubah data dari satu bentuk atau format ke bentuk atau format lain untuk mencapai tujuan analisis atau kebutuhan bisnis tertentu. Transformasi data melibatkan manipulasi dan restrukturisasi data guna meningkatkan kualitas, konsistensi, dan relevansi informasi. Tujuan utama dari transformasi data adalah untuk membuat data lebih sesuai dengan kebutuhan analisis atau pemrosesan lebih lanjut, seperti pemodelan statistik, visualisasi data, atau aplikasi bisnis. Proses ini mencakup berbagai teknik, termasuk normalisasi data untuk membawa data ke skala yang seragam, agregasi data untuk menyusun data ke dalam tampilan ringkasan, dan pengubahan tipe data untuk mengubah jenis nilai data, serta banyak lagi. transformasi data memainkan peran penting dalam persiapan data yang baik sebelum analisis, memastikan bahwa data siap digunakan dengan benar dan memberikan wawasan yang berharga.

Selain itu, transformasi data juga dapat melibatkan penggabungan data dari berbagai sumber yang berbeda, sehingga data yang terpisah dapat diintegrasikan menjadi satu kesatuan yang konsisten. Proses penggabungan ini memastikan bahwa data yang berasal dari sistem atau departemen yang berbeda dapat digunakan bersama untuk analisis lintas fungsi atau menyatukan informasi tentang suatu entitas tertentu. Dengan mengintegrasikan dan mentransformasi data, organisasi dapat menghindari duplikasi, memperbaiki ketidakakuratan, dan menciptakan satu

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



sumber kebenaran (*single source of truth*) yang menyediakan data yang terpercaya dan konsisten untuk berbagai keperluan.

Transformasi data juga digunakan untuk menyajikan data dalam format yang lebih mudah dimengerti dan relevan bagi para pemangku kepentingan. Sebagai contoh, data mentah yang kompleks atau berantakan dapat diubah menjadi format yang lebih singkat, seperti grafik atau tabel ringkasan, agar lebih mudah diinterpretasikan dan dianalisis. Penggunaan teknik transformasi seperti pengategorian, pengdiskritisasi, dan penyederhanaan data dapat mempermudah pemahaman terhadap pola dan tren dalam data yang kompleks. Dengan menggunakan *data transformation* secara bijaksana, organisasi dapat memaksimalkan potensi data yang mereka miliki dan mendapatkan wawasan yang lebih mendalam, yang pada akhirnya dapat membantu dalam pengambilan keputusan yang lebih akurat dan efektif.

2.13.3 Feature Engineering

Feature Engineering adalah proses penting dalam menganalisis data dan membangun model prediktif di mana fitur atau variabel yang relevan dan informatif diekstraksi, diciptakan, atau diubah dari data mentah untuk meningkatkan kinerja model. *Feature Engineering* adalah proses pemilihan, manipulasi, dan transformasi data menal menjadi fitur yang dapat digunakan dalam *supervised learning* (Harshil, 2021). Fitur merupakan representasi dari informasi yang terkandung dalam data dan menjadi elemen kunci dalam proses pemodelan. Proses *Feature Engineering* melibatkan pemahaman mendalam tentang data dan domain masalah, sehingga fitur yang paling relevan dan berdampak signifikan terhadap hasil prediksi dapat diidentifikasi. *Feature Engineering* mencakup menciptakan fitur berdasarkan kombinasi fitur lain, mengubah data, mengurangi dimensi, atau menggunakan pengetahuan domain untuk meningkatkan representasi data. Fitur yang baik membantu model untuk menemukan pola yang lebih kompleks dan menyajikan data dengan cara yang lebih berarti, sehingga meningkatkan akurasi dan generalisasi model.

Selain itu, *Feature Engineering* juga berperan dalam mengatasi masalah data yang hilang atau *outlier*, sehingga memperkuat keandalan model. Sebagai

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



contoh, teknik pengisian dapat digunakan untuk mengisi nilai yang hilang pada fitur tertentu berdasarkan pola dari fitur lainnya. Pengenalan fitur khusus yang menggambarkan adanya data yang tidak biasa atau ekstrim dapat membantu model untuk memahami variasi ekstrem yang ada dalam data. Dengan memilih dan menggabungkan fitur yang tepat, *Feature Engineering* memungkinkan model untuk belajar dan menggeneralisasi lebih baik dari data yang tersedia, sehingga meningkatkan kinerja model dalam mengambil keputusan atau membuat prediksi.

Salah satu aspek penting dari *Feature Engineering* adalah pengawasan dan evaluasi yang berkelanjutan untuk memastikan fitur-fitur yang digunakan tetap relevan dan efektif. Kinerja model dapat dipengaruhi oleh fitur-fitur yang digunakan, dan dengan adanya perubahan data atau perubahan kebutuhan bisnis, fitur-fitur yang paling informatif dapat berubah. Oleh karena itu, *Feature Engineering* merupakan proses yang terus-menerus ditingkatkan seiring dengan evolusi data dan perubahan dalam domain masalah. Dengan pendekatan yang teliti dan kreatif dalam *Feature Engineering*, analisis data dan model prediktif dapat menjadi lebih kuat dan mampu memberikan wawasan berharga untuk pengambilan keputusan dan penyelesaian masalah yang lebih baik.

2.13.4 Data Reduction

Data reduction atau reduksi data, adalah proses mengurangi volume atau dimensi data dengan mempertahankan sebagian besar informasi penting, sehingga memungkinkan analisis dan pemrosesan data menjadi lebih efisien dan efektif (Javatpoint, 2022) (Flin, 2020). Tujuan utama dari pengurangan data adalah untuk mengatasi masalah yang muncul akibat data yang sangat besar, kompleks, atau berlebihan. Proses ini melibatkan teknik-teknik seperti mengurangi dimensi, memilih fitur, dan sampling data. Mengurangi dimensi melibatkan memproyeksikan data ke dimensi yang lebih rendah tetapi mempertahankan sebagian besar informasi penting. Memilih fitur, di sisi lain, memilih *subset* fitur yang paling relevan dan informatif dari data mentah. Sementara itu, sampling data melibatkan mengambil sampel representatif dari data asli untuk mewakili keseluruhan populasi. Dengan mengurangi volume data yang harus diproses, waktu dan sumber daya yang diperlukan untuk analisis dan pemodelan dapat berkurang, sehingga mempercepat proses dan mengurangi beban pemrosesan.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Namun, penting untuk diingat bahwa pengurangan data harus dilakukan dengan hati-hati untuk meminimalkan risiko kehilangan informasi atau menyebabkan distorsi dalam analisis. Reduksi data yang berlebihan dapat mengakibatkan hilangnya informasi penting dan mengarah pada hasil analisis yang tidak akurat atau tidak relevan. Oleh karena itu, memilih teknik Pengurangan Data yang tepat sangat penting untuk memastikan bahwa data yang dikurangi masih mencerminkan fitur penting dan representatif dari data asli. Selain itu, evaluasi kinerja setelah proses pengurangan data diterapkan juga diperlukan untuk memastikan bahwa hasil analisis tetap valid dan relevan.

Pengurangan data memainkan peran penting dalam pengelolaan data besar dan rumit, terutama dalam bidang seperti analisis Big Data, *machine learning*, dan *data mining*. Dengan menggunakan teknik-teknik pengurangan data yang tepat, organisasi dapat mengatasi tantangan dalam memproses, menyimpan, dan mengelola data yang besar dan beragam. Dengan data yang lebih terfokus dan efisien, organisasi dapat mendapatkan wawasan yang lebih cepat, mengoptimalkan kinerja model dan analisis, serta meningkatkan kecepatan dalam pengambilan keputusan berdasarkan informasi yang relevan.

2.13.5 Data Normalization

Data normalization atau normalisasi data merupakan proses perubahan nilai-nilai data mentah menjadi bentuk yang nantinya dapat diolah dan dianalisis dengan lebih efektif dan efisien. Tujuan utama dari proses ini, yaitu untuk menghilangkan redundansi data dan memberikan standar informasi untuk alur kerja yang lebih baik (Melisa, 2023). Normalisasi data dibutuhkan apabila terdapat atribut-atribut dari *dataset* yang memiliki skala atau nilai rentang yang berbeda. Ketika dalam sebuah *dataset* terdapat banyak karakteristik, tetapi nilainya bervariasi atau atribut pada skala yang berbeda, nantinya model yang dibangun dapat menghasilkan prediksi yang tidak akurat saat melakukan proses *data mining*. Oleh sebab itu, perlu dilakukan langkah normalisasi data.

Pada normalisasi data terdapat beberapa metode dalam pelaksanaannya, berikut di antaranya:

- Normalisasi *Z-Score*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Normalisasi *Z-Score* atau yang biasa dikenal dengan *standard score* merupakan sebuah teknik normalisasi dengan berdasarkan pada mean dan standar deviasi. Teknik ini mentransformasikan data dari nilai skala umum dengan angka mean sama dengan nol dan standar deviasi adalah satu. Dengan teknik ini, akan bermanfaat dalam menganalisis data yang memerlukan perbandingan nilai dengan nilai rata-rata. Pada *Z-Score* nilai baru yang dihasilkan berdasarkan dengan perbedaan yang dimiliki dari nilai rata – rata dan juga nilai standar deviasi (Whendasmoro dan Joseph, 2022) (Trivusi, 2022).

$$X_{stand} = \frac{x - \text{mean}(x)}{\text{std}_{dev}(x)} \quad (2.7)$$

$$\text{std}_{dev}(x) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x - \text{mean}(x))^2} \quad (2.8)$$

- Normalisasi Min-Max

Teknik normalisasi min-max digunakan untuk mengubah nilai-nilai data ke dalam rentang nilai antara 0 hingga 1. Data asli mengalami modifikasi linear dalam prosedur normalisasi data ini. Setiap nilai diubah dengan menggunakan rumus berikut (Trivusi, 2022):

$$X_{norm} = \frac{x' - \min(x)}{\max(x) - \min(x)} (\text{new}_{max}(x) - \text{new}_{min}(x)) + \text{new}_{min}(x) \quad (2.9)$$

Keterangan:

- X adalah atribut data
- Min(x) dan max(x) adalah nilai absolut minimal dan maksimal dari x
- X' adalah nilai lama dari setiap entri data
- New_max(x) dan new_min(x) adalah nilai min dan max dari range



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

- Normalisasi *Decimal Scaling*
Teknik normalisasi ini bekerja dengan cara mengubah nilai-nilai data ke dalam rentang nilai antara -1 dan 1. Teknik ini memanfaatkan konsep pergeseran desimal dari bilangan bulat yang diterapkan pada setiap nilai data. Pada prosesnya, teknik ini menggunakan rumus seperti berikut (Trivusi, 2022):

$$v' = \frac{v}{10^j} \quad (2.10)$$

Keterangan:

- V' adalah nilai baru setelah dilakukan penykalaan desimal
- V merupakan nilai atribut
- J adalah bilangan bulat terkecil sehingga $\max(|v'|) < 1$

2.14 Resampling Data

Resampling data adalah langkah khusus yang dilakukan dalam fase pra-pemrosesan *machine learning*. Fase pra-pemrosesan menyiapkan data sebelum digunakan untuk melatih dan menguji model *machine learning*. Langkah *resampling data* terutama dilakukan untuk memecahkan masalah ketidakseimbangan kelas dalam *dataset*. Ketidakseimbangan kelas terjadi ketika ada lebih banyak contoh dari satu kelas (kelas mayoritas) daripada contoh dari kelas lain (kelas minoritas). Masalah ini biasa terjadi pada tugas klasifikasi di mana kelas yang diprediksi jarang terjadi. Pada *resampling data* terbagi menjadi tiga teknik, yaitu *Oversampling*, *Undersampling*, dan *Combination*. Tiga teknik tersebut dijelaskan sebagai berikut.

2.14.1 Oversampling

Oversampling adalah teknik pengolahan data untuk mengatasi masalah ketidakseimbangan kelas dalam kumpulan data. Seperti namanya, ini adalah teknik pemilihan acak kelas minoritas dan menduplikasinya untuk menambahkan data kelas minoritas (Akalbir, 2021). Ketidakseimbangan kelas terjadi ketika ada lebih banyak contoh dari satu kelas (kelas mayoritas) daripada contoh dari kelas lain



(kelas minoritas). Ini sering terjadi dalam tugas klasifikasi di mana kelas yang ingin Anda prediksi jarang terjadi.

Oversampling mereplikasi sampel dari kategori minoritas sehingga jumlahnya sama dengan jumlah sampel dari kategori mayoritas. Dengan demikian, *oversampling* meningkatkan jumlah contoh di kelas minoritas, memungkinkan model untuk belajar lebih baik dari kelas yang kurang terwakili. Teknik *oversampling* dapat dilakukan dengan menduplikasi contoh kelas minoritas, atau menggunakan teknik seperti SMOTE (*Synthetic Minority Over-sampling Technique*) untuk membuat contoh sintetik berdasarkan data yang ada.

Meskipun *oversampling* dapat membantu masalah ketidakseimbangan kelas, ada juga beberapa potensi masalah yang harus diperhatikan. Peningkatan jumlah contoh kelas minoritas dapat menyebabkan *overfitting*, di mana model terlalu fokus pada kelas minoritas dan tidak menggeneralisasi dengan baik data yang sebelumnya tidak terlihat. Oleh karena itu, keakuratan mengelola dan memvalidasi *dataset oversampling* penting untuk memastikan bahwa model yang dihasilkan adalah model yang andal dan secara efektif menyelesaikan masalah ketidakseimbangan kelas.

2.14.2 *Undersampling*

Undersampling adalah teknik pengolahan data untuk mengatasi masalah ketidakseimbangan kelas dalam kumpulan data. Ketidakseimbangan kelas terjadi ketika ada lebih banyak contoh dari satu kelas (kelas mayoritas) daripada contoh dari kelas lain (kelas minoritas). *Undersampling* menghilangkan sampel dari kelas mayoritas sehingga jumlahnya sama dengan jumlah sampel dari kelas minoritas. Dengan cara ini, *undersampling* mengurangi jumlah contoh di kelas mayoritas, memungkinkan model untuk lebih fokus pada kelas minoritas yang kurang terwakili.

Meskipun *undersampling* dapat membantu memecahkan masalah ketidakseimbangan kelas, ada beberapa pertimbangan. Mengurangi jumlah contoh di kelas mayoritas dapat menyebabkan hilangnya informasi yang serius, terutama bila jumlah informasi aslinya sudah terbatas. Selain itu, *undersampling* dapat membuat model menjadi tidak stabil dan tidak dapat diandalkan jika jumlah contoh

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

di kelas mayoritas terlalu sedikit. Oleh karena itu, akurasi dalam mengatur *dataset* dan mengevaluasi kinerja model setelah *subsampling* adalah kunci untuk mendapatkan hasil yang baik dan efektif dalam menyelesaikan masalah ketidakseimbangan kelas.

2.14.3 Combination Sampling

Combination sampling adalah teknik *resampling* data yang menggabungkan dua atau lebih metode *resampling* untuk memecahkan masalah ketidakseimbangan kelas dalam kumpulan data. Tujuan dari teknik ini adalah untuk memanfaatkan setiap metode dan mengatasi kelemahannya. Dengan menggunakan kombinasi teknik *resampling*, maka dapat membuat sampel sintetis yang lebih beragam dan mencerminkan distribusi data asli secara dekat.

2.15 SMOTEENN

SMOTEENN (SMOTE-ENN) adalah teknik *resampling* data yang menggabungkan dua pendekatan, yaitu SMOTE (*Synthetic Minority Over-sampling Technique*) dan ENN (*Edited Nearest Neighbors*). Oleh karena itu, SMOTEENN termasuk dalam teknik *resampling* gabungan atau kombinasi. SMOTE-ENN efektif dalam mengatasi ketidakseimbangan kelas dalam prediksi pelanggan churn, terutama ketika jumlah pelanggan churn jauh lebih kecil daripada jumlah pelanggan tidak churn dalam *dataset* (Chaudhary *et al.*, 2023).

Pertama, SMOTE cenderung *oversample* kelas minoritas dengan menghasilkan sampel sintetis baru berdasarkan sampel yang ada. Ini membantu meningkatkan jumlah contoh di kelas minoritas, memungkinkan model untuk belajar lebih baik dari kelas yang kurang terwakili. Namun, *oversampling* juga dapat menghasilkan sampel sintetis yang *noise* dan menurunkan generalisasi model dalam data baru yang belum pernah ada sebelumnya. ENN kemudian digunakan untuk meng-*undersample* kelas mayoritas dengan menghilangkan beberapa sampel yang dianggap *outlier* atau sampel yang tidak jelas. Teknik ini membantu mengurangi jumlah contoh di kelas mayoritas, memungkinkan model untuk lebih fokus pada kelas minoritas yang kurang terwakili. Dengan menggabungkan SMOTE dan ENN, SMOTEENN berupaya mengatasi masalah kemacetan sintetis



yang dapat terjadi dengan *oversampling* sederhana dan memitigasi potensi masalah dengan menghilangkan sampel *fuzzy* atau *noise* dari kelas mayoritas.

2.16 *User Acceptance Test (UAT)*

User Acceptance Testing (UAT) merupakan proses verifikasi bahwa solusi yang dibuat dalam sistem sudah sesuai untuk pengguna (Suprpto, 2021). *User Acceptance Testing (UAT)* dilakukan untuk mengukur tingkat penerimaan sistem oleh pengguna (B.Firdaus, I. Rokhmawati, dan A. Wicaksono, 2021). Bentuk kuesioner UAT bisa beragam tergantung pada kebutuhan dan tujuan pengujian sistem yang sedang dilakukan. Namun, umumnya kuesioner tersebut akan memuat pertanyaan-pertanyaan yang mengajak pengguna untuk memberikan penilaian terhadap berbagai aspek sistem, seperti antarmuka, fungsionalitas, kemudahan penggunaan, dan performa sistem. Pertanyaan-pertanyaan dalam survei tersebut mungkin akan menitikberatkan pada pendapat pengguna tentang faktor-faktor yang bisa mempengaruhi sejauh mana mereka bersedia menerima perubahan dalam cara mereka melakukan analisis bisnis. UAT memiliki tujuan untuk menghimpun masukan langsung dari pengguna akhir. Hal ini bertujuan untuk memastikan bahwa sistem yang telah dikembangkan sesuai dengan harapan dan kebutuhan pengguna, sehingga meningkatkan tingkat kepuasan mereka. UAT membantu dalam mengidentifikasi masalah atau kekurangan yang mungkin tidak terdeteksi selama tahap pengembangan sebelumnya. Pengguna akhir memiliki kemampuan untuk menemukan masalah yang mungkin tidak terlihat oleh tim pengembang. Dengan melibatkan pengguna dalam pengujian akhir, risiko terjadinya kesalahan atau masalah di lingkungan produksi dapat ditekan karena masalah dapat diatasi sebelum sistem digunakan secara luas.

2.17 *Contract Acceptance Testing (CAT)*

Contract Acceptance Testing (CAT) adalah tahap pengujian yang terjadi setelah perangkat lunak atau sistem telah dikembangkan dan diuji oleh tim pengembang. Pada tahap ini, perangkat lunak atau sistem diserahkan kepada pelanggan atau klien untuk diuji dan diverifikasi sesuai dengan persyaratan kontrak atau kesepakatan yang telah dibuat. Tujuan dari CAT ini untuk memastikan bahwa perangkat lunak atau sistem yang dikembangkan telah memenuhi persyaratan dan harapan pelanggan sesuai dengan kesepakatan yang telah ditetapkan sebelumnya.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Uji ini bertujuan untuk memverifikasi bahwa perangkat lunak bekerja sesuai dengan tujuan bisnis dan memenuhi semua persyaratan fungsional dan nonfungsional yang disepakati. Dalam beberapa kasus, *Contract Acceptance Testing* dapat memiliki kesamaan dengan *User Acceptance Testing* (UAT). CAT merupakan faktor penentu dalam kepuasan kontrak antara vendor perangkat lunak dan pelanggan (Hsia *et al.*, 2002).



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

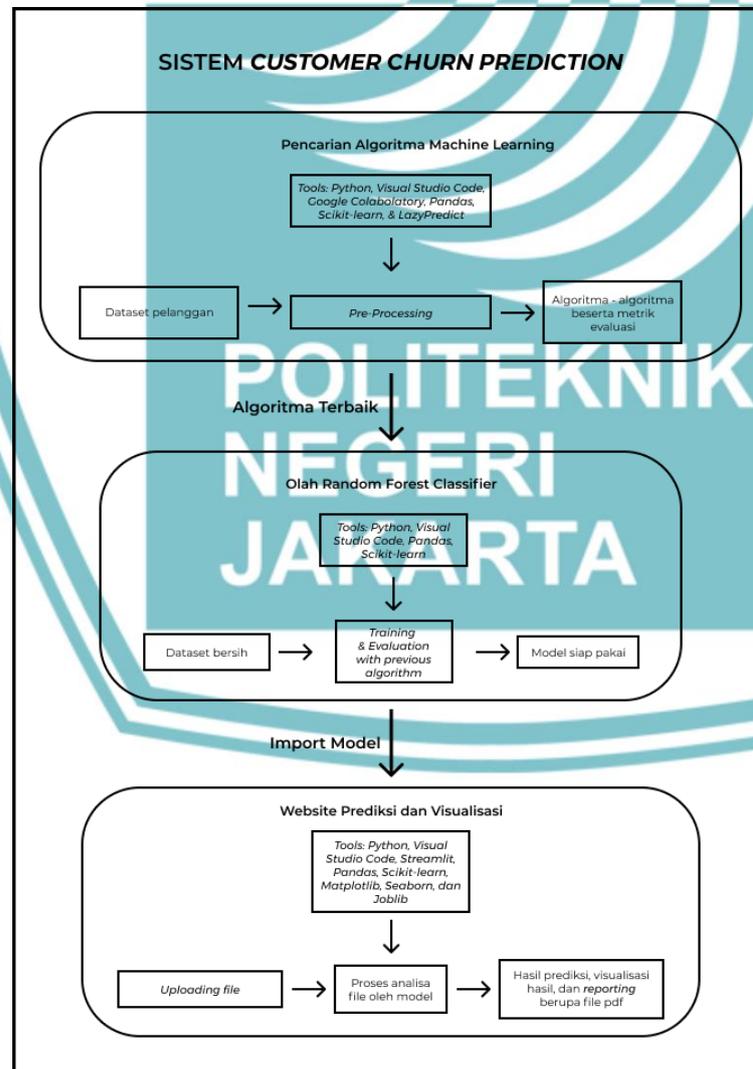


- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

BAB III PERANCANGAN DAN REALISASI

Pemanfaatan teknologi komputasi, khususnya *machine learning*, sangat bermanfaat bagi seluruh sektor perindustrian. Ranah telekomunikasi juga tidak luput dari kebermanfaatannya teknologi *machine learning*. Dengannya, perusahaan dapat menganalisis kebiasaan dari suatu pelanggan. Dengan analisis tersebut, nantinya dapat dibuat sebuah prediksi mengenai loyalitas dari pelanggan tersebut. Oleh karena itu, dibuatlah suatu perancangan mengenai sistem prediksi loyalitas pelanggan dengan memanfaatkan kemajuan teknologi, *machine learning*.

3.1 Perancangan Tugas Akhir



Gambar 3.1 Diagram Sistem Customer Churn Prediction



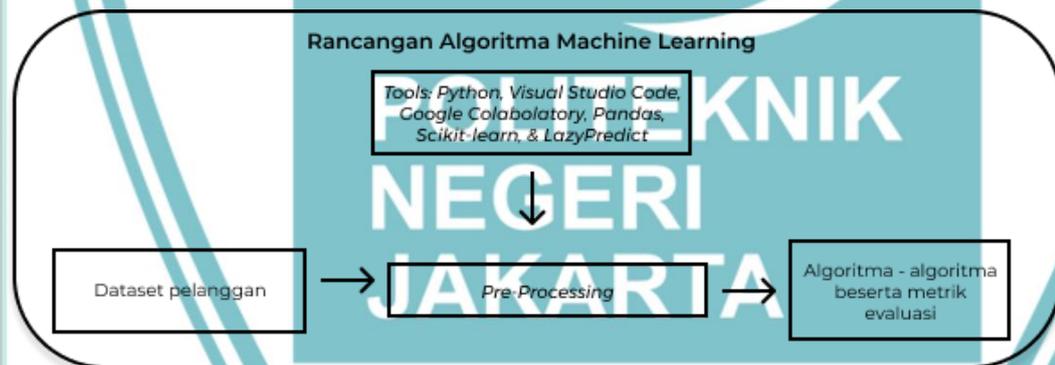
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Dalam melakukan perancangan Sistem Customer Churn Prediction diperlukan adanya perancangan sistem terlebih dahulu. Perancangan tersebut terbagi menjadi 3 bagian, yaitu rancangan algoritma *machine learning*, rancangan model *Random Forest Classifier*, dan rancangan *website* prediksi dan visualisasi.

Pada perancangan algoritma *machine learning* meliputi deskripsi algoritma *machine learning*, spesifikasi algoritma *machine learning*, diagram blok algoritma *machine learning*, arsitektur algoritma *machine learning*, dan cara kerja algoritma *machine learning*. Pada perancangan model *Random Forest Classifier* meliputi deskripsi model *Random Forest Classifier*, spesifikasi model *Random Forest Classifier*, diagram blok model *Random Forest Classifier*, arsitektur model *Random Forest Classifier*, dan cara kerja model *Random Forest Classifier*. Pada perancangan *website* prediksi dan visualisasi meliputi deskripsi *website* prediksi dan visualisasi, spesifikasi *website* prediksi dan visualisasi, diagram blok *website* prediksi dan visualisasi, arsitektur *website* prediksi dan visualisasi, dan cara kerja *website* prediksi dan visualisasi.

3.1.1 Rancangan Algoritma *Machine Learning*



Gambar 3.2 Rancangan Algoritma *Machine Learning*

Gambar 3.2 merupakan rancangan dari pencarian algoritma *machine learning*. Pada gambar tersebut dijelaskan peta kerja dari proses pencarian algoritma *machine learning* terbaik. Algoritma yang dipilih nantinya akan diterapkan pada tahap berikutnya, yaitu pembuatan model *machine learning*.

3.1.1.1 Deskripsi Algoritma *Machine Learning*

Model *machine learning* yang akan dibuat nantinya akan dibantu pencariannya dengan menggunakan *library* LazyPredict. LazyPredict sangat



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

membantu karena *library* ini membandingkan berbagai model *machine learning* secara cepat dan efisien. Nantinya, keluaran yang dihasilkan oleh LazyPredict dapat digunakan untuk memahami bagaimana setiap model bekerja dengan *dataset* yang diberikan dan dapat ditentukan model mana yang paling tepat untuk digunakan. Berdasarkan keluaran yang dihasilkan oleh LazyPredict, selanjutnya akan dianalisis untuk mengevaluasi model dengan melihat metrik evaluasi.

3.1.1.2 Spesifikasi Pencarian Algoritma *Machine Learning*

Spesifikasi algoritma *machine learning* akan ditampilkan pada Tabel 3.1.

Tabel 3.1 Spesifikasi Untuk Pencarian Model *Machine Learning*

No	Nama	Spesifikasi
1	Python	Versi 3.9.6
2	Visual Studio Code	Versi 1.79.1
3	Google Colabulatory	Versi Update 2023/06/02
4	Pandas	Versi 1.5.3
5	Scikit-Learn	Versi 1.2.2
6	LazyPredict	Versi 0.2.12
7	Imbalanced-learn	Versi 0.10.1

Pada Tabel 3.1 di atas merupakan daftar spesifikasi yang akan digunakan dalam mendukung pencarian algoritma *machine learning*. Python digunakan sebagai bahasa pemrograman dalam merancang algoritma *machine learning*. Visual Studio Code digunakan sebagai IDE (*Integrated Development Environment*) selama proses perancangan algoritma *machine learning* dilakukan. Google Colabulatory digunakan sebagai penunjang dari Visual Studio Code. Pandas sebagai *library* yang digunakan untuk pemrosesan and analisis data yang akan berguna dalam pre-processing data dan selama proses perancangan algoritma *machine learning*. Scikit-Learn sebagai *machine learning library* yang digunakan untuk penugasan-penugasan tertentu, seperti klasifikasi, regresi, pengelompokkan, pemrosesan data, pengurangan dimensi, pemilihan fitur, dan evaluasi model. LazyPredict sebagai *library* yang digunakan untuk menyederhanakan dan mempercepat proses eksplorasi dasar pada saat pemodelan *machine learning*.



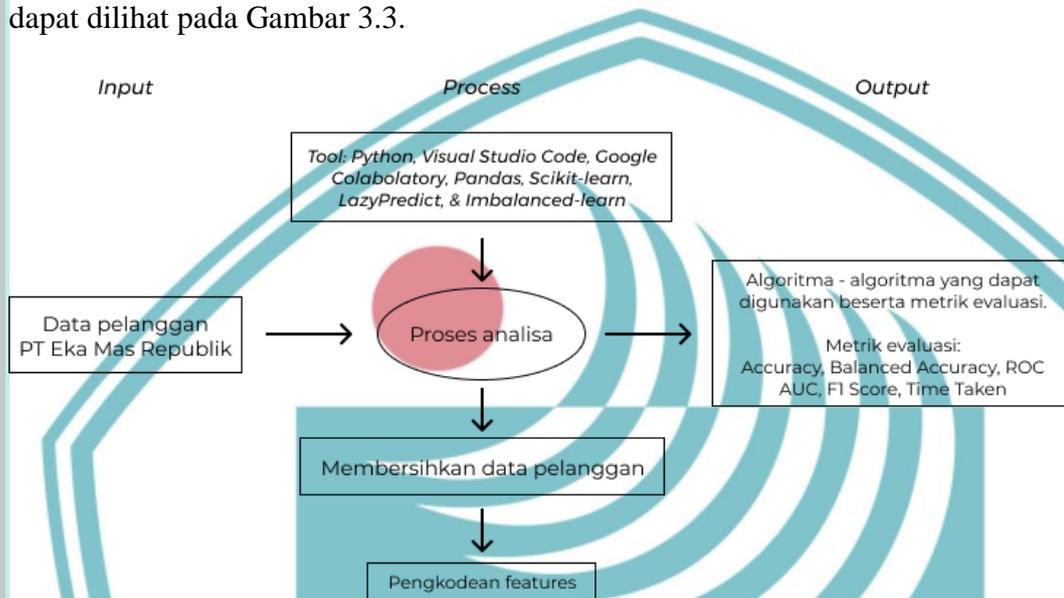
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Imbalanced-learn sebagai *library* yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam *machine learning*.

3.1.1.3 Diagram Blok Perancangan Algoritma *Machine Learning*

Diagram blok dalam melakukan pencarian algoritma *machine learning* dapat dilihat pada Gambar 3.3.



Gambar 3.3 Diagram Blok Perancangan Algoritma *Machine Learning*

Tahapan untuk mencari algoritma *machine learning* yang nantinya akan digunakan pada sistem prediksi churn, meliputi:

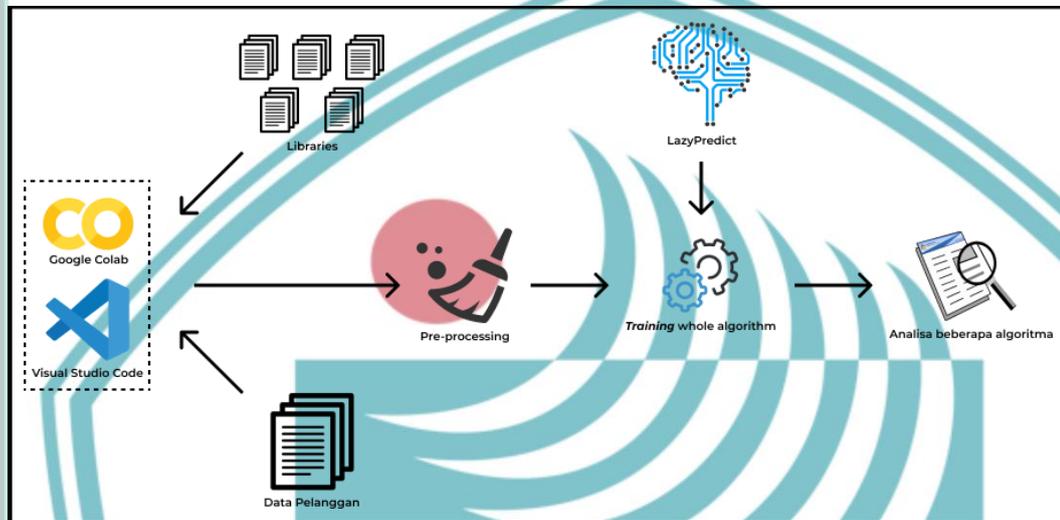
1. Melakukan pencarian data pelanggan, meliputi data profil demografis dari pelanggan MyRepublik. Lokasi pengambilan data dilakukan di Tangerang, Banten, Indonesia.
2. Melakukan proses analisis terhadap *dataset* yang telah dikumpulkan. Proses analisis berupa pembersihan *dataset*. *Dataset* ini akan terdiri atas 10 fitur yang terbagi menjadi 4 fitur kategorikal dan 6 fitur numerikal. Fitur-fitur ini antara lain, *Area Name*, *Plan*, *Tv Plan*, *Advance Promo*, *Complaint by Customer Service*, *Complaint by Email*, *Complaint by Media Social*, *Complaint by Telegram*, *Complaint by Whatsapp*, dan *Complaint by WIC*.
3. Pada tahap proses pembersihan data pelanggan terbagi menjadi 3 bagian, yaitu seleksi *feature*, cek data duplikasi, dan cek data *null*.
4. Setelah *dataset* telah terorganisir dan bersih, selanjutnya data-data yang berupa kategorikal akan melalui proses pengodean data. Pengodean data ini



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

dilakukan agar algoritma *machine learning* yang di-generate nantinya dapat mempelajarinya dengan baik. Data kategorikal biasanya akan dikodekan dengan cara mengubah *value* kategorikal menjadi suatu nilai numerik yang mewakili dari data kategorikal tersebut.

3.1.1.4 Arsitektur Perancangan Algoritma *Machine Learning*



Gambar 3.4 Arsitektur Perancangan Algoritma *Machine Learning*

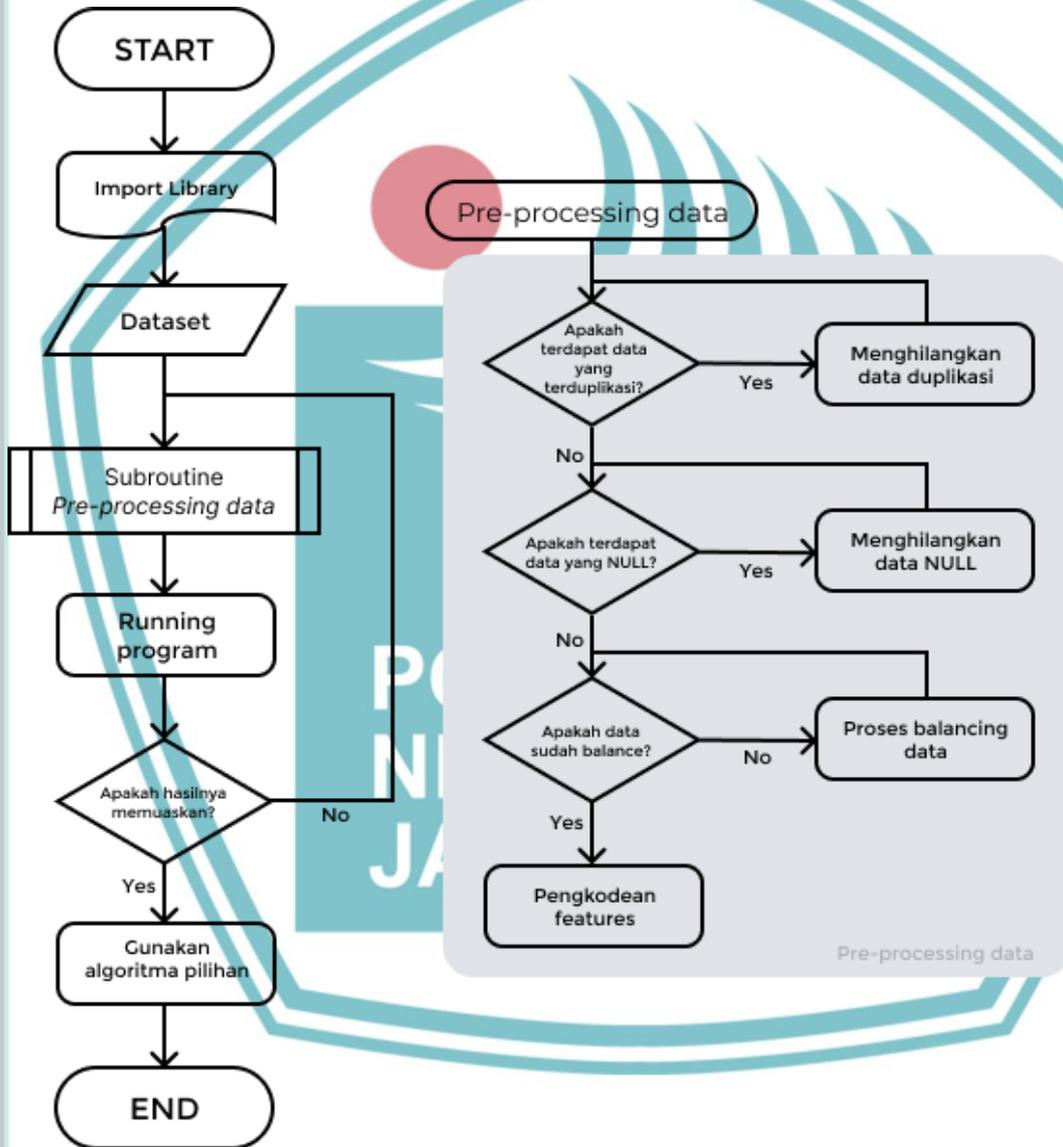
Gambar 3.4 merupakan arsitektur dari perancangan algoritma *machine learning*. Perancangan akan dimulai dengan mempersiapkan platform untuk *coding*, yaitu Google Colab dan Visual Studio Code. Untuk platform tersebut, digunakan Google Colab Versi 1.79.1 dan Visual Studio Code Versi Update 2023/06/02. Setelah itu disiapkan *library-library* yang diperlukan, serta siapkan juga *dataset* yang berisi data *demographics profile*, *usage*, dan *customer services activity* pelanggan. Data *demographics profile*, yaitu data yang menggambarkan pengelompokan demografis atau sebuah segmen pasar dan informasi demografis berisi kemungkinan-kemungkinan perilaku pelanggan. Data *usage* di sini diartikan sebagai data yang mendeskripsikan kepada penggunaan pelanggan. Data *customer services activity* diartikan sebagai data yang mendeskripsikan jumlah terjadinya gangguan dan komplain.

Setelah persiapan platform dan *library* selesai, proses akan berlanjut ke *pre-processing data*. Setelah data pelanggan melalui proses *pre-processing* akan menghasilkan sebuah *dataset* yang bersih dan siap digunakan untuk proses selanjutnya. *Dataset* yang telah bersih akan dilatih dengan *library* yang bernama



LazyPredict. *Library* ini menjadi otak pada proses pencarian algoritma *machine learning*. *Library* ini akan menghasilkan *output* berupa rangkuman dari algoritma-algoritma *classifier* dari algoritma-algoritma yang ada di luaran sana. Selanjutnya, sekumpulan algoritma-algoritma tersebut dianalisis berdasarkan hasil metrik evaluasi yang terdapat pada masing-masing algoritma.

3.1.1.5 Cara Kerja Algoritma *Machine Learning*



Gambar 3.5 Flowchart Pencarian Algoritma Machine Learning

Gambar 3.5 merupakan *flowchart* dari cara kerja pencarian algoritma *machine learning*, yang dijelaskan lebih detail pada berikut ini:

- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



© Hak Cipta milik Politeknik Negeri Jakarta

1. *Import Library*

Import library yang dibutuhkan selama proses yang dibutuhkan oleh Python. Di sini akan digunakan *library* Pandas, Scikit-Learn, dan LazyPredict.

2. *Dataset*

Dataset selanjutnya dimasukkan dan dibaca oleh *library* Pandas. Data ini dapat berupa data *demographics profile* (menggambarkan pengelompokan demografis atau sebuah segmen pasar dan informasi demografis berisi kemungkinan-kemungkinan perilaku pelanggan), *usage* (data yang mendeskripsikan kepada penggunaan pelanggan), dan *customer services activity* (data yang mendeskripsikan jumlah terjadinya gangguan dan komplain). Data-data yang telah terkumpul selanjutnya akan dihimpun menjadi satu, dan ini dapat disebut dengan *dataset*. Untuk data yang sudah terhimpun meliputi data a CID, ID, Area Network, Cluster, *Plan*, *Tv Plan*, Promo Advance, Complaint by User, *Complaint by Email*, Complaint by Media Social, *Complaint by Telegram*, Complaint by WhatsApp, *Complaint by WIC*. Data ini mencakup data akhir tahun 2022 hingga data awal tahun 2023.

3. *Pre-processing Data*

Pada *pre-processing data* terdapat beberapa hal yang harus ditinjau, ditindaklanjuti, dan dilakukan. Berikut beberapa hal yang perlu dicermati:

- *Data Duplikasi*

Memastikan tidak adanya data yang mengalami duplikasi merupakan hal yang wajib dilakukan. Data yang terduplikasi dapat merusak kualitas dari *dataset* itu sendiri. Hal ini dapat menyebabkan kesalahan dalam analisis statis, dapat menyebabkan *overfitting*, distorsi pada analisis keterkaitan, peningkatan waktu dan sumber daya yang dibutuhkan, dan dapat menyebabkan bias dalam hasil analisis dan kesimpulan yang diambil.

- *Data Null*

Kehadiran *null* data pada suatu *dataset* sangat berpengaruh pada kualitas data. Dengan tidak adanya sebuah data, maka data akan mengalami kehilangan

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



© Hak Cipta milik Politeknik Negeri Jakarta

suatu informasi. Data *null* juga dapat menyebabkan kesulitan dalam melakukan analisis statistik yang memerlukan data lengkap, seperti perhitungan nilai *mean*, standar deviasi, atau melakukan uji hipotesis. Kehilangan data juga akan berpengaruh pada pengambilan keputusan. Efek terbesarnya, yaitu akan berpengaruh pada pemodelan dan prediksi serta akan terjadi bias dalam analisis.

- Keseimbangan Data

Pada saat hendak melakukan pencarian algoritma terbaik keseimbangan data akan sangat mempengaruhi distribusi dari data itu sendiri, yang nantinya akan berdampak ke kualitas data pada saat dipelajari oleh algoritma-algoritma. Ketika terjadi ketidakseimbangan data antara kelas mayoritas (not-churn) dan kelas minoritas (churn), model akan cenderung mempelajari untuk memprediksi kelas mayoritas lebih baik dan kelas minoritas akan terabaikan. Ini mengakibatkan model nantinya menjadi bias dan tidak efektif dalam memprediksi kelas minoritas.

- Pengodean Fitur

Pengodean ini penting dikarenakan model *machine learning* pada umumnya hanya dapat bekerja dengan data numerik. Oleh karena itu, fitur – fitur yang bersifat kategorikal harus mengalami proses *encoding* atau pengodean terlebih dahulu. *Encoding* ini penting juga untuk menghindari masalah dimensi. Jika fitur – fitur kategorikal tidak *encoded*, maka setiap kategori akan menjadi fitur baru yang terpisah. Ini dapat menyebabkan masalah dimensi atau "*curse of dimensionality*" di mana jumlah fitur menjadi sangat besar dibandingkan dengan jumlah sampel. Hal ini dapat mengakibatkan kinerja model yang buruk dan waktu komputasi yang lama. Dengan *encoding* fitur kategorikal menjadi representasi numerik, jumlah fitur dapat dikurangi dan memperbaiki efisiensi model.

4. Importing *Library* LazyPredict

Dalam proses pencarian algoritma yang paling sesuai dengan keadaan *dataset* yang dimiliki akan digunakan sebuah *library* yang bernama LazyPredict. Oleh karena itu, Langkah awal yang harus dilakukan, yaitu *install* dan melakukan inisiasi *import* LazyPredict dalam sintaks yang dibuat.

5. Running Program

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

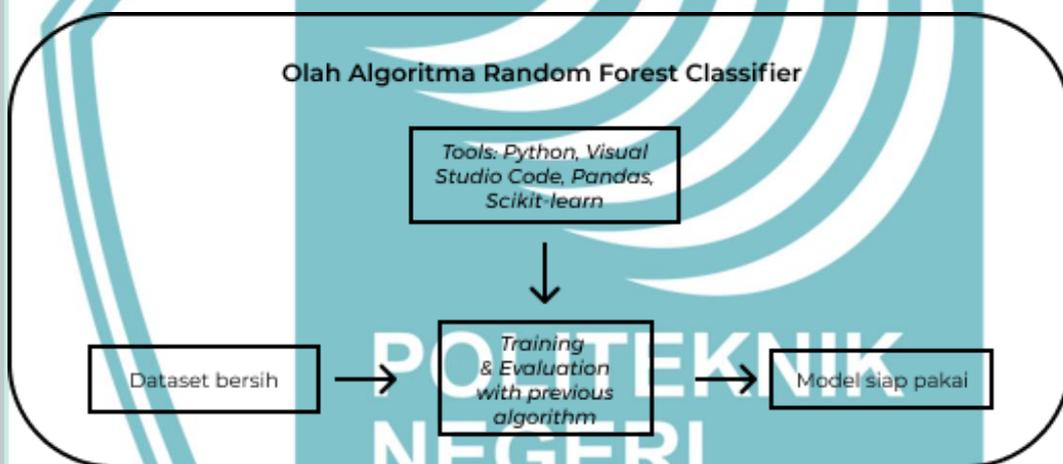
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Dengan *library* LazyPredict kumpulan algoritma-algoritma yang ada di luar sana tidak perlu dilakukan satu per satu. Hanya dengan beberapa baris sintaks kode, LazyPredict akan melakukan otomatisasi proses *training* dan *testing* dengan seluruh algoritma yang tersedia di dalamnya.

6. Fiksasi Algoritma

Setelah serangkaian proses di atas dilakukan, selanjutnya kumpulan algoritma yang telah LazyPredict latih ini akan dianalisis. Analisis meliputi, *accuracy*, *balanced accuracy*, *ROC AUC*, *F1 Score*, dan *time taken*. Setelah dianalisis, selanjutnya akan diambil satu algoritma yang memiliki metrik evaluasi terbaik.

3.1.2 Perancangan Model *Machine Learning*



Gambar 3.6 Perancangan Model Random Forest Classifier

Berdasarkan proses pencarian algoritma-algoritma *machine learning*, *Random Forest Classifier* adalah salah satu algoritma yang terbaik. Oleh karena itu, pada perancangan kali ini akan menggunakan algoritma *Random Forest Classifier* dalam pembuatan model *machine learning*.

3.1.2.1 Deskripsi Model *Machine Learning*

Sistem *customer churn prediction* yang dibuat ini berfungsi untuk mengidentifikasi pelanggan yang berpotensi untuk berhenti berlangganan pada perusahaan penyedia layanan internet dan multimedia, MyRepublik. Sistem ini didesain berdasarkan model *machine learning*. Model di sini dipilih *Random Forest Classifier* karena kemampuannya yang baik dalam menganalisis dan memproses



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

dataset. Persiapan data dengan baik adalah kunci dari hasil model yang maksimal. Pengolahan data dilakukan dengan pra-pemrosesan data diantaranya, pembersihan data, imputasi data yang hilang, dan normalisasi fitur. Hasil keluaran tidak hanya berupa prediksi churn dari seorang pelanggan, tetapi juga akan memberikan *insight* penting mengenai faktor-faktor yang paling berpengaruh terhadap keputusan churn pelanggan. Pengetahuan akan hal ini sangat berharga untuk perusahaan dalam menyusun strategi retensi pelanggan lebih efektif dan tepat sasaran.

3.1.2.2 Spesifikasi Model *Machine Learning*

Spesifikasi model *Machine Learning* akan ditampilkan pada Tabel 3.2.

Tabel 3.2 Spesifikasi Model *Machine Learning*

No	Nama	Spesifikasi
1	Python	Versi 3.9.6
2	Visual Studio Code	Versi 1.79.1
3	Google Colabulatory	Versi Update 2023/06/02
4	Pandas	Versi 1.5.3
5	Scikit-Learn	Versi 1.2.2
6	Imbalanced-learn	Versi 0.10.1

Pada Tabel 3.2 di atas merupakan daftar spesifikasi yang akan digunakan dalam mendukung perancangan algoritma *machine learning*. Python digunakan sebagai bahasa pemrograman dalam merancang algoritma *machine learning*. Visual Studio Code digunakan sebagai IDE (*Integrated Development Environment*) selama proses perancangan algoritma *machine learning* dilakukan. Google Colabulatory digunakan sebagai penunjang dari Visual Studio Code. Pandas sebagai *library* yang digunakan untuk pemrosesan and analisis data yang akan berguna dalam *pre-processing data* dan selama proses perancangan algoritma *machine learning*. Scikit-Learn sebagai *machine learning library* yang digunakan untuk penugasan-penugasan tertentu, seperti klasifikasi, regresi, pengelompokkan, pemrosesan data, pengurangan dimensi, pemilihan fitur, dan evaluasi model. Imbalanced-learn sebagai *library* yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam *machine learning*.

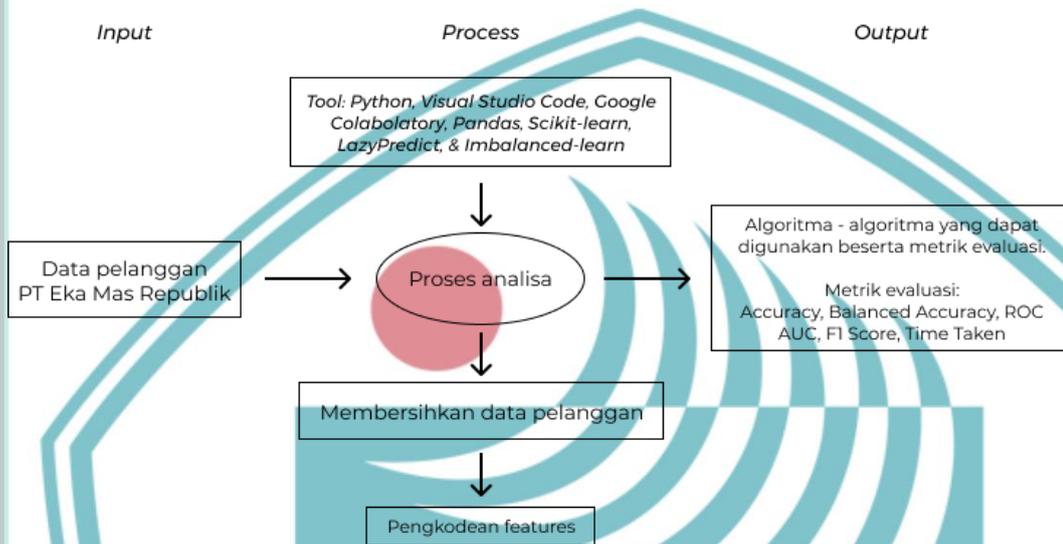


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

3.1.2.3 Diagram Blok Model *Machine Learning*

Adapun diagram blok dari model *machine learning* dapat dilihat pada Gambar 3.7.



Gambar 3.7 Diagram Blok Model *Machine Learning*

Tahapan untuk merancang sistem *customer churn prediction*, meliputi:

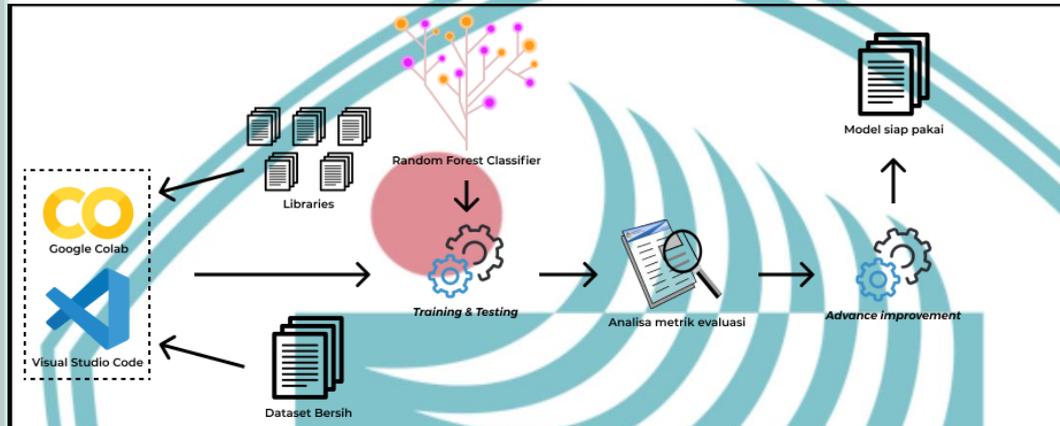
1. Input pada tahap ini, yaitu *dataset* yang telah melalui pra-proses data. *Dataset* ini sudah dalam kondisi bersih dan siap pakai.
2. *Dataset* ini selanjutnya akan melalui proses *training & testing* dengan algoritma *machine learning Random Forest Classifier*.
3. Setelah didapatkan hasil dari proses *training & testing*, selanjutnya model akan dianalisis berdasarkan metrik evaluasinya. Metrik evaluasi di antaranya, *accuracy, f1 score, recall*, dan *ROC AUC*.
4. Hasil dari Analisa metrik-metrik evaluasi ini nantinya akan sebagai tolak ukur apakah model yang dihasilkan telah sesuai dengan kebutuhan yang dimaksudkan.
5. Jika hasil dari Analisa ini dirasa kurang memuaskan, maka ada beberapa tahapan lanjut yang diharapkan dengan menambahkan beberapa tahapan lanjut ini, dapat menghasilkan model yang jauh lebih baik daripada model sebelumnya.



6. *Output* dari proses di atas, yaitu menghasilkan model dengan tingkat performansi baik, dapat dijalankan, dan diterapkan pada *website* prediksi churn pelanggan nantinya.

3.1.2.4 Arsitektur Model *Machine Learning*

Adapun arsitektur untuk model *machine learning* berbentuk seperti berikut.



Gambar 3.8 Arsitektur Model *Machine Learning*

Gambar 3.8 merupakan arsitektur dari pembuatan model *machine learning*. Perancangan akan dimulai dengan mempersiapkan platform untuk *coding*, yaitu Google Colab dan Visual Studio Code. Untuk platform tersebut, digunakan Google Colab Versi 1.79.1 dan Visual Studio Code Versi Update 2023/06/02. Setelah itu disiapkan *library-library* yang diperlukan, serta siapkan *dataset* bersih yang telah melewati pra-proses data sebelumnya.

Setelah setup *library* dan *dataset* bersih, langkah selanjutnya, yaitu proses *training & testing*. Proses ini dilakukan dengan menggunakan algoritma *Random Forest Classifier*. Algoritma ini merupakan algoritma *tree-based*. Setelah melalui proses *training & testing*, *output* dari algoritma ini akan menampilkan metrik evaluasi yang meliputi, *Accuracy Score*, *F1 Score*, *Recall Score*, dan *ROC AUC Score*. Jika nilai dari masing-masing metrik evaluasi belum sudah sesuai dan tidak cukup untuk menggambarkan model yang diinginkan, maka akan melalui *improvement* lebih lanjut. Jika dirasa nilai dari masing-masing metrik evaluasi sudah sesuai dan cukup untuk menggambarkan model yang diinginkan, maka model sudah siap digunakan dan diimplementasikan pada *website* prediksi churn pelanggan.

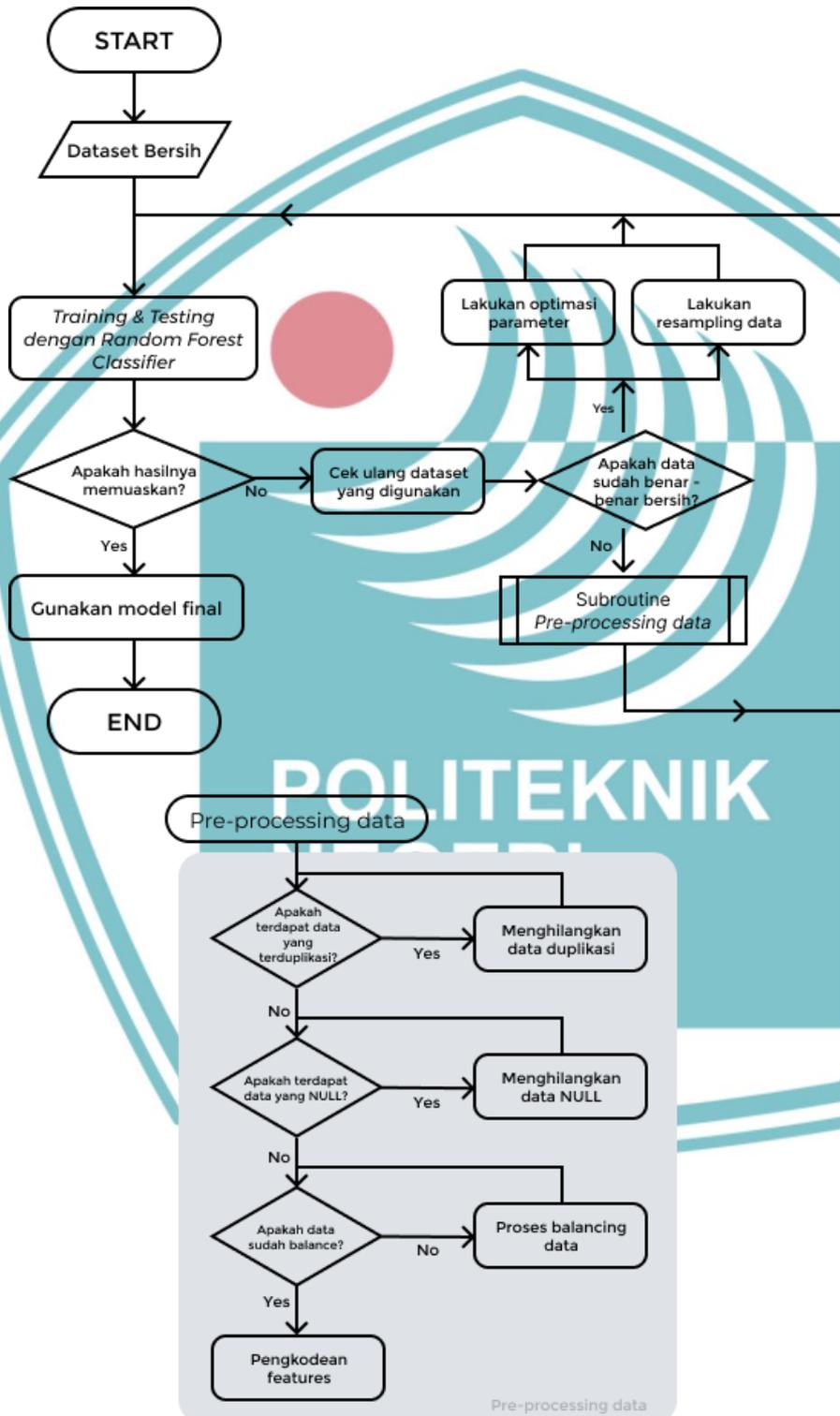
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



3.1.2.5 Cara Kerja Model *Machine Learning*

Cara kerja sistem prediksi churn dilakukan dengan tahapan seperti yang tercantum pada *flowchart* yang ditunjukkan pada Gambar 3.9.



Gambar 3.9 Flowchart Cara Kerja Model Machine Learning

- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



© Hak Cipta milik Politeknik Negeri Jakarta

Dari Gambar 3.9 menunjukkan cara kerja dari model *Machine Learning* yang dibuat, penjelasan lebih rinci dapat dijelaskan di bawah ini.

1) *Dataset* Bersih

Pada proses sebelumnya, pencarian algoritma, *pre-processing data* telah dilewati. Oleh karena itu, di proses ini langsung menggunakan *dataset* yang siap digunakan.

2) *Training & Testing*

Pada tahap ini, *dataset* yang bersih akan dilakukan *training & testing* dengan menggunakan algoritma *Random Forest Classifier*. Di sini *dataset* akan dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. *Dataset* akan mengalami pengacakan terlebih dahulu, setelah itu akan dibagi dengan komposisi 80% untuk data *training* dan 20% untuk data *testing*. Rasio untuk pembagian *dataset* ini tidak boleh dilakukan secara asal. Data *training* yang berjumlah lebih sedikit dibandingkan data *testing* dapat membuat model menjadi *overfitting*, ketidakseimbangan kelas, dan lainnya. Begitu pula sebaliknya, ada juga dampak dari kurangnya data *testing*. Dampak ini mulai dari evaluasi yang tidak akurat hingga ketidakseimbangan dalam pengambilan keputusan.

3) Tingkat Kepuasan

Setelah dilakukan *training & testing* dengan menggunakan algoritma *Random Forest Classifier*, selanjutnya metrik-metrik evaluasi akan dianalisis. Metrik-metrik evaluasi ini meliputi *accuracy*, *f1 score*, *recall*, dan *ROC AUC*. Jika dirasa metrik-metrik evaluasi ini sudah cukup untuk menggambarkan model yang diinginkan, maka proses berhenti sampai titik ini. Jika dirasa kurang, maka ada beberapa tahap lanjutan. Dimulai dari pengecekan ulang *dataset* yang digunakan, apakah data tersebut sudah bersih dan siap digunakan atau data tersebut ternyata masih belum siap pakai. Jika didapati bahwa *dataset* yang digunakan masih belum siap pakai, maka harus melewati pra-proses data lagi. Dimulai dari pengecekan dan penanganan data duplikat, *null*, dan keseimbangan data, hingga bermuara pada pengodean *features*. Berbeda kondisi jika *dataset* sudah diyakini bersih dan terbukti siap pakai, maka akan melalui dua tahap lanjutan, yaitu optimasi parameter

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

dan *resampling data*. Optimasi parameter ini mengacu pada proses pencarian kombinasi parameter yang optimal untuk model yang digunakan. Optimasi ini bertujuan untuk meningkatkan kinerja dari model berdasarkan metrik-metrik evaluasi. Sementara itu, proses *resampling data* itu bermaksud memanipulasi ulang atau mengubah distribusi data yang ada. Hal ini bertujuan untuk mengatasi masalah ketidakseimbangan kelas atau distribusi yang tidak merata antar kelas dalam *dataset*. Setelah semua dilakukan, proses akan kembali kepada *training* dan *testing* model, hingga menemukan racikan yang terbaik.

4) Final Model

Pada saat racikan terbaik telah ditemukan, model dapat digunakan dan diimplementasikan pada *website* prediksi churn pelanggan.

3.1.3 Perancangan Website Prediksi Churn Pelanggan

3.1.3.1 Deskripsi Website Prediksi Churn Pelanggan

Website prediksi churn pelanggan ini dibuat agar perusahaan lebih memahami dan dapat merespon kebutuhan dan keinginan pelanggan mereka. Dengan memahami faktor-faktor yang menjadi pendorong churn, perusahaan dapat merancang dan menerapkan strategi retensi pelanggan yang efektif. Website prediksi churn juga dapat mengoptimalkan sumber daya perusahaan. Dengan mengetahui pelanggan mana yang berpotensi churn, perusahaan dapat memfokuskan pada Upaya retensi untuk segmen tersebut dan ini akan menghindari pengeluaran yang tidak perlu pada pelanggan yang tidak mungkin akan churn. Dalam jangka panjang, *website* ini dapat membantu perusahaan membangun hubungan lebih kuat dan berkelanjutan dengan pelanggan, meningkatkan loyalitas merek, dan memperpanjang umur pelanggan.

Website ini memiliki dua fitur utama, yaitu fitur prediksi tunggal dan fitur prediksi *batch*. Pada fitur prediksi tunggal, nantinya *user* akan memasukkan nilai dari setiap variabel seorang pelanggan. Input ini meliputi, *Area Name*, *Plan*, *Tv Plan*, *Advance Promo*, *Complaint by Customer Service*, *Complaint by Email*, *Complaint by Social Media*, *Complaint by Telegram*, *Complaint by Whatsapp*, dan *Complaint by WIC*. Di sini *user* dapat menginput variabel – variabel tersebut sesuai dengan pelanggan mana yang ingin dia prediksi. Fitur ke dua, yaitu prediksi *batch*.



Di sini *user* akan memprediksi data banyak pelanggan sekaligus dan disatukan dalam sebuah *file batch* dengan ekstensi *.csv*. Website ini juga memiliki halaman yang memiliki konten berupa deskripsi singkat tentang *website* ini dan sedikit informasi tentang penulis.

3.1.3.2 Spesifikasi Website Prediksi Churn Pelanggan

Spesifikasi *website* prediksi churn akan ditampilkan pada Tabel 3.3.

Tabel 3.3 Spesifikasi Website Prediksi Churn Pelanggan

No	Nama	Spesifikasi
1	Python	Versi 3.9.6
2	Visual Studio Code	Versi 1.79.1
3	Google Colabatory	Versi Update 2023/06/02
4	Streamlit	Versi 1.23.1
5	Pandas	Versi 1.5.3
6	Joblib	Versi 1.2.0
7	Matplotlib	Versi 3.7.1
8	Seaborn	Versi 1.2.2
9	PyPDF2	Versi 3.0.1

Python digunakan sebagai bahasa pemrograman dalam merancang algoritma *machine learning*. Visual Studio Code digunakan sebagai IDE (*Integrated Development Environment*) selama proses perancangan algoritma *machine learning* dilakukan. Google Colabatory digunakan sebagai penunjang dari Visual Studio Code. Streamlit sebagai sebuah *library open-source* dari bahasa pemrograman python yang berguna pemuatan *interface* aplikasi *data science*. Pandas sebagai *library* yang digunakan untuk pemrosesan and analisis data yang akan berguna dalam pre-processing data dan selama proses perancangan algoritma *machine learning*. Joblib sebagai *library* Python yang berguna untuk mempermudah penyimpanan dan pembacaan objek Python yang besar. Matplotlib, *library* Python yang digunakan untuk visualisasi data dan pengolahan *file* PDF. Seaborn, *library* yang digunakan untuk visualisasi data. PyPDF2, *lirary* Python yang digunakan untuk proses pembuatan *churn report* ke dalam bentuk PDF.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

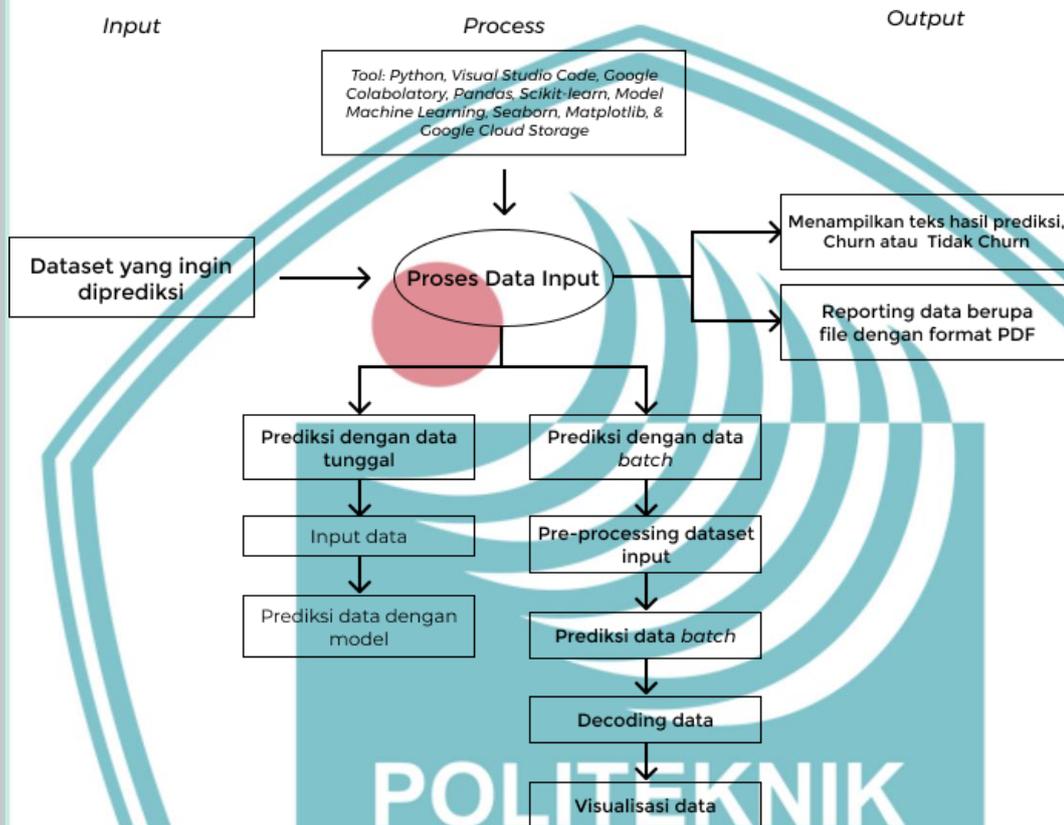


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

3.1.3.3 Blok Diagram Website Prediksi Churn Pelanggan

Adapun diagram blok dari *website* prediksi churn pelanggan dapat dilihat pada Gambar 3.10.



Gambar 3.10 Blok Diagram Website Prediksi Churn Pelanggan

Tahapan untuk merancang sistem customer churn prediction, meliputi:

1. *Input data*. Di sini *website* disediakan dua fitur, yaitu prediksi tunggal dan prediksi *batch*. Jika *user* ingin melakukan prediksi tunggal, maka *user* hanya butuh *input data* yang sesuai dengan data pelanggan secara manual pada *website*. Data yang diinputkan pada *website*, antara lain *Area Name*, *Plan*, *Tv Plan*, *Advance Promo*, *Complaint by Customer Service*, *Complaint by Email*, *Complaint by Social Media*, *Complaint by Telegram*, *Complaint by Whatsapp*, dan *Complaint by WIC*. Jika *user* ingin melakukan prediksi dengan data *batch*, *user* harus memasukkan data dengan format *file .csv*.
2. Melakukan proses analisis dan prediksi terhadap data *input* dari *user*.



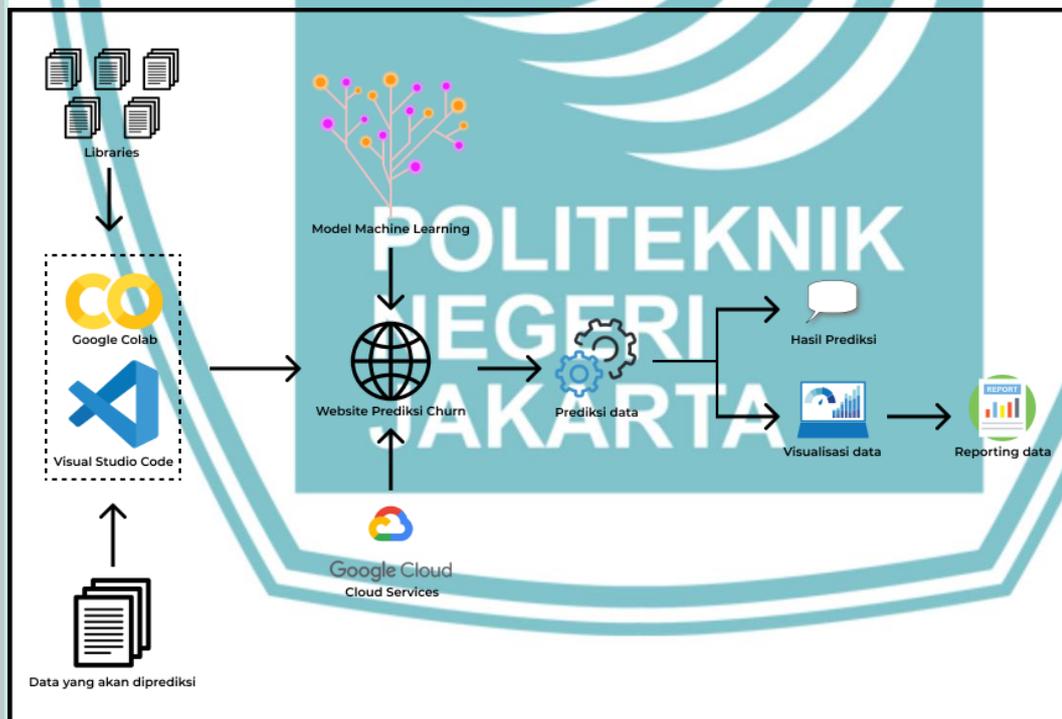
© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

3. Jika dilakukan proses prediksi data tunggal, maka data langsung akan diprediksi pada *website* dan *outputnya* nanti akan berupa pesan bahwa sesuai data tersebut, apakah pelanggan akan churn atau tidak.
4. Jika dilakukan proses prediksi data *batch*, maka data mengalami beberapa tahap terlebih dahulu sebelum hasil data tersebut ditampilkan. Pertama, data tersebut akan melalui proses pengodean terlebih dahulu agar dapat dibaca oleh model *machine learning* yang telah dibuat. Selanjutnya data yang telah melalui tahap pengodean, data akan diprediksi. Hasil data prediksi nanti selanjutnya akan disatukan dengan data semula. Selanjutnya data final akan diproses dengan *decoding*, di sini data akan dikembalikan kepada nilai aslinya supaya nanti pada saat visualisasi data dapat dibaca dengan mudah.
5. Terakhir, data *batch* yang telah diproses, akan divisualisasikan pada *website*. Selain itu, hasilnya juga akan dibuat *reporting file* dengan format PDF.

3.1.3.4 Arsitektur Website Prediksi Churn Pelanggan



Gambar 3.11 Gambar Arsitektur Website Prediksi Churn Pelanggan

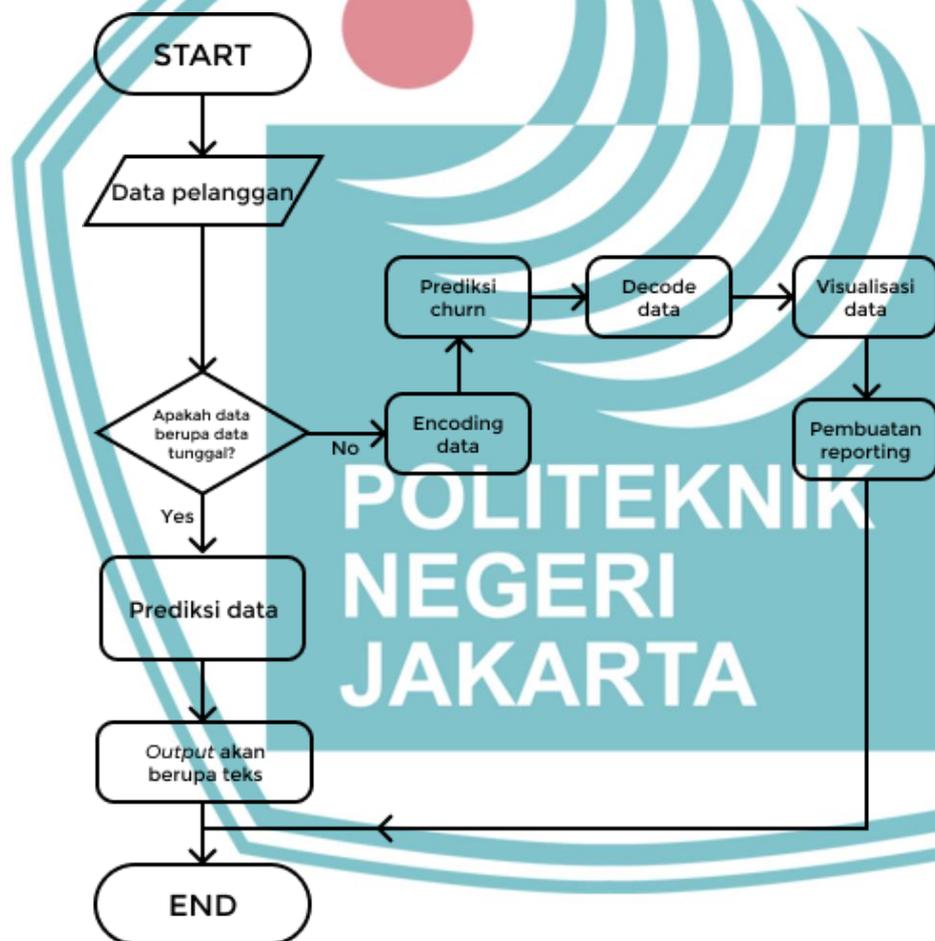
Gambar 3.11 merupakan arsitektur dari *website* prediksi churn pelanggan. Dimulai dengan *import library* yang dibutuhkan. Website yang dibuat akan memuat sebuah model *machine learning* yang telah dibuat sebelumnya. Website ini *running* pada sebuah *cloud platform*, Google Cloud. Setelah persiapan *website* selesai,



selanjutnya proses prediksi data. Di sini dibagi menjadi dua bagian, yaitu data tunggal dan data *batch*. *Output* dari prediksi data tunggal hanya akan berupa teks yang menampilkan data tersebut akan churn atau tidak. Sementara itu, untuk data *batch* akan ditampilkan hasil prediksi berupa grafik-grafik yang merepresentasikan data tersebut. Selanjutnya hasil visualisasi tersebut akan dibuat menjadi *reporting file* dalam format PDF.

3.1.3.5 Cara Kerja Website Prediksi Churn Pelanggan

Cara kerja *website* prediksi churn pelanggan dilakukan dengan tahapan yang tercantum pada *flowchart* yang ditunjukkan pada Gambar 3.8.



Gambar 3.12 *Flowchart* Cara Kerja Website Prediksi Churn Pelanggan

Dari Gambar 3.12 menunjukkan cara kerja *website* prediksi churn pelanggan yang dibuat, penjelasan lebih rinci dapat dijelaskan di bawah ini.

- 1) Data Pelanggan

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

User akan dihadapi dengan pilihan jenis dua data, yaitu data tunggal dan data *batch*. User akan memilih sesuai dengan kebutuhan.

2) Prediksi Data

Data yang telah diinput, akan diprediksi. Jika *user* melakukan prediksi dengan data tunggal, maka data langsung diprediksi. Jika *user* melakukan prediksi dengan data *batch*, maka data akan melalui proses pengodean data terlebih dahulu. Setelah data terkodekan dengan baik, maka data tersebut akan diprediksi. Setelah diprediksi data akan mengalami proses *decoding*. Di sini data akan dikembalikan ke dalam nilai aslinya. Setelah nilai data terkembalikan dengan baik, maka data siap divisualisasikan.

3) Visualisasi dan *Reporting*

Data *batch* yang telah diprediksi dan melalui proses *decode*, data siap divisualisasikan. Visualisasi data akan merepresentasikan data yang bersangkutan dengan hasil churn data tersebut.

3.2 Realisasi Tugas Akhir

3.2.1 Realisasi Algoritma *Machine Learning*

Tahap ini dilakukan realisasi dari algoritma *machine learning*. Hasil dari tahap ini akan menetapkan algoritma mana yang nantinya akan diproses dan menjadi model *machine learning* yang akan digunakan. Realisasi algoritma *machine learning* akan menggunakan bahasa Python sebagai bahasa pemrograman dan Google Colaboratory sebagai *text editor*, seperti pada Tabel 3.1.

3.2.1.1 Pre-processing Data

Realisasi algoritma *machine learning* dimulai dengan proses *pre-processing data*. Proses *pre-processing data* akan dilaksanakan dengan *flowchart* cara kerja pencarian algoritma *machine learning* pada Gambar 3.5.

1) Melakukan instalasi dan *import library*

Langkah pertama yang harus dilakukan adalah *install* dan *import* semua *library* yang dibutuhkan. Berikut prosesnya.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```

!pip install lazypredict

Collecting lazypredict
  Downloading lazypredict-0.2.12-py2.py3-none-any.whl (12 kB)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from lazypredict) (8.1.6)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from lazypredict) (1.2.2)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from lazypredict) (1.5.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from lazypredict) (4.65.0)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from lazypredict) (1.3.1)
Requirement already satisfied: lightgbm in /usr/local/lib/python3.10/dist-packages (from lazypredict) (3.3.5)
Requirement already satisfied: xgboost in /usr/local/lib/python3.10/dist-packages (from lazypredict) (1.7.6)
Requirement already satisfied: wheel in /usr/local/lib/python3.10/dist-packages (from lightgbm->lazypredict) (0.41.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from lightgbm->lazypredict) (1.23.5)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from lightgbm->lazypredict) (1.10.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->lazypredict) (3.2.0)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas->lazypredict) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->lazypredict) (2022.7.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil->2.8.1->pandas->lazypredict) (1.16.0)
Installing collected packages: lazypredict
Successfully installed lazypredict-0.2.12

[4] import pandas as pd
import joblib
import lazypredict
import seaborn as sns

from lazypredict.Supervised import LazyClassifier
from imblearn.combine import SMOTEENN, SMOTETomek
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import LabelEncoder
from imblearn.combine import SMOTETENN
from sklearn.model_selection import train_test_split
from matplotlib import pyplot as plt
from sklearn.metrics import classification_report, roc_curve, accuracy_score, f1_score, roc_auc_score, precision_score, recall_score

```

Gambar 3.13 Library Installation & Importing

2) Melakukan *pre-processing data*

Pada langkah ini dilakukan pengecekan data duplikat, *null*, dan keseimbangan data. Selain itu, pada langkah ini juga dilakukan penghilangan data duplikasi dan data *null*. Dalam proses menyeimbangkan data, dilakukan dengan cara *resampling data*. *Resampling data* dilakukan dengan menggunakan *library* SMOTE-TOMEK.

POLITEKNIK
NEGERI
JAKARTA



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
# Load the dataset
df = pd.read_csv("aslibaruinimah.csv")

null_counts = df.isnull().sum()
df.drop_duplicates(inplace=True)

plt.title("Churn Count")

#Resampling Data
#Area Name
unique_area = df['Area Name'].unique()
value_area = range(len(unique_area))
dict_area = dict(zip(unique_area, value_area))
df['Area Name'] = df['Area Name'].replace(dict_area)

#Plan
unique_plan = df['Plan'].unique()
value_plan = range(len(unique_plan))
dict_plan= dict(zip(unique_plan, value_plan))
df['Plan'] = df['Plan'].replace(dict_plan)

#TV Plan
unique_tvplan = df['Tv Plan'].unique()
value_tvplan = range(len(unique_tvplan))
dict_tvplan= dict(zip(unique_tvplan, value_tvplan))
df['Tv Plan'] = df['Tv Plan'].replace(dict_tvplan)

#Advance Promo
unique_promo = df['Advance Promo'].unique()
value_promo = range(len(unique_promo))
dict_promo= dict(zip(unique_promo, value_promo))
df['Advance Promo'] = df['Advance Promo'].replace(dict_promo)

# Memisahkan fitur (X) dan target (y)
X = df.drop('Churn', axis=1)
y = df['Churn']

# Membagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Menerapkan SMOTE-Tomek pada data latih
smote_tomek = SMOTETomek()
X_train_resampled, y_train_resampled = smote_tomek.fit_resample(X_train, y_train)

# Membuat DataFrame baru dengan hasil resampling
df_smotetomek = pd.DataFrame(X_train_resampled, columns=X_train.columns)
df_smotetomek['Churn'] = y_train_resampled

df_smotetomek.to_csv('smotetomek.csv', index=False)
```

Gambar 3.14 Pre-processing Data



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

3.2.1.2 LazyPredict

Setelah *pre-processing data* dilakukan, selanjutnya dilakukan proses untuk *training* dan *testing* dengan menggunakan *library* LazyPredict.

```
df = pd.read_csv('smotetomek.csv')

X = df.drop('Churn', axis=1)
y = df['Churn']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

clf = LazyClassifier(predictions=True)
models, predictions = clf.fit(X_train, X_test, y_train, y_test)

print(models)
```

Gambar 3.15 Generate Algoritma Machine Learning

Setelah dilakukan *training* dan *testing* data dengan menggunakan LazyPredict, maka akan muncul hasilnya seperti berikut.

Model	Accuracy	ROC	AUC	Time Tak
ExtraTreesClassifier	NaN	0.83	0.83	23.98
RandomForestClassifier	NaN	0.83	0.83	26.76
BaggingClassifier	NaN	0.83	0.83	4.38
DecisionTreeClassifier	NaN	0.81	0.81	0.

Gambar 3.16 Hasil *Training & Testing* Menggunakan LazyPredict

Gambar 3.16 merupakan hasil dari *training* dan *testing* data dengan menggunakan *library* LazyPredict. Hasilnya menampilkan algoritma-algoritma yang masuk ke dalam jenis Classifier. Berdasarkan data tersebut, diambil algoritma *Random Forest Classifier*. Hal ini karena *Random Forest Classifier* salah satu algoritma yang memiliki nilai metrik evaluasi terbaik di antara banyaknya pilihan algoritma yang di-generate LazyPredict.

3.2.2 Realisasi Model *Machine Learning*

3.2.2.1 Dataset

Dataset yang telah dibuat, 'smotetomek.csv' masih dirasa kurang optimal untuk model *machine learning* yang akan dibuat. Oleh karena itu, *dataset* awal akan melalui *resampling data* ulang dengan menggunakan metode berbeda. Di sini akan menggunakan *resampling data* dengan nama SMOTEENN.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

1) Import *library*

Langkah pertama yang harus dilakukan adalah *import* semua *library* yang dibutuhkan. Berikut prosesnya.

```
[4] import pandas as pd
import joblib
import lazypredict
import seaborn as sns

from lazypredict.Supervised import LazyClassifier
from imblearn.combine import SMOTEENN, SMOTETomek
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import LabelEncoder
from imblearn.combine import SMOTEENN
from sklearn.model_selection import train_test_split
from matplotlib import pyplot as plt
from sklearn.metrics import classification_report, roc_curve, accuracy_score, f1_score, roc_auc_score, precision_score, recall_score
```

Gambar 3.17 Import Library

Proses selanjutnya, yaitu *pre-processing data* dengan menggunakan *resampling data* SMOTEENN.

2) *Pre-processing data*

Pada langkah ini dilakukan pengecekan data duplikat, *null*, dan keseimbangan data. Selain itu, pada langkah ini juga dilakukan penghilangan data duplikasi dan data *null*. Dalam proses menyeimbangkan data, dilakukan dengan cara *resampling data*. *Resampling data* dilakukan dengan menggunakan *library* SMOTEENN.

POLITEKNIK
NEGERI
JAKARTA



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
df = pd.read_csv("aslibaruinimah.csv")

null_counts = df.isnull().sum()
df.drop_duplicates(inplace=True)

churn_count = sns.catplot(x = 'Churn', kind='count', palette='magma', data = df, height = 6)
for ax in churn_count.axes.flat:
    ax.bar_label(ax.containers[0])
plt.title("Churn Count")

#Resampling Data
#Area Name
unique_area = df['Area Name'].unique()
value_area = range(len(unique_area))
dict_area = dict(zip(unique_area, value_area))
df['Area Name'] = df['Area Name'].replace(dict_area)

#Plan
unique_plan = df['Plan'].unique()
value_plan = range(len(unique_plan))
dict_plan= dict(zip(unique_plan, value_plan))
df['Plan'] = df['Plan'].replace(dict_plan)

#TV Plan
unique_tvplan = df['Tv Plan'].unique()
value_tvplan = range(len(unique_tvplan))
dict_tvplan= dict(zip(unique_tvplan, value_tvplan))
df['Tv Plan'] = df['Tv Plan'].replace(dict_tvplan)

#Advance Promo
unique_promo = df['Advance Promo'].unique()
value_promo = range(len(unique_promo))
dict_promo= dict(zip(unique_promo, value_promo))
df['Advance Promo'] = df['Advance Promo'].replace(dict_promo)

# Memisahkan fitur (X) dan target (y)
X = df.drop('Churn', axis=1)
y = df['Churn']

# Membagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Memisahkan fitur (X) dan target (y)
X = df.drop('Churn', axis=1)
y = df['Churn']

# Menerapkan SMOTEENN
smote_enn = SMOTEENN()
X_resampled, y_resampled = smote_enn.fit_resample(X, y)

# Menggabungkan kembali fitur dan target yang sudah diresampling
df_resampled = pd.concat([pd.DataFrame(X_resampled, columns=X.columns), pd.Series(y_resampled, name='Churn')], axis=1)

df_resampled.to_csv('smoteenn.csv', index=False)
```

Gambar 3.18 Resampling Data Dengan SMOTEENN

3.2.2.2 Training dan Testing

Setelah *library* dan *dataset* dimasukkan, maka selanjutnya yaitu proses *training* dan *testing dataset* tersebut dengan menggunakan algoritma *Random Forest Classifier*.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```

# Memisahkan fitur (X) dan target (y)
X = df_resampled.drop('Churn', axis=1)
y = df_resampled['Churn']

# Membagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Membuat objek Random Forest Classifier
rf_model = RandomForestClassifier()

# Melatih model menggunakan data latih
rf_model.fit(X_train, y_train)

# Melakukan prediksi pada data uji
y_pred = rf_model.predict(X_test)

# Menghitung akurasi prediksi
accuracy = accuracy_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
roc = roc_auc_score(y_test, y_pred)
prec = precision_score(y_test, y_pred)

print("Accuracy:", accuracy)
print("F1 Score:", f1)
print("Recall Score:", recall)
print("ROC AUC:", roc)
print("Precision Score:", prec)

```

Gambar 3.19 Training & Testing Dengan Resampled Data

Pada proses di atas, dilakukan pemisahan fitur dan data target, proses *splitting data* dengan rasio 80:20 antara *training data* : *testing data*. Setelah melakukan proses *training* dan *testing*, akan ditampilkan skor metrik evaluasi. Berikut metrik evaluasi hasil *training & testing* dengan menggunakan SMOTEENN.

```

Accuracy: 0.948586424907551
F1 Score: 0.9471533085380309
Recall Score: 0.9733702956989247
ROC AUC: 0.9498410557392322
Precision Score: 0.9223115497890632

```

Gambar 3.20 Hasil Metrik Evaluasi SMOTEENN

Berdasarkan 3.20, nilai dari metrik evaluasi tersebut selanjutnya akan dianalisis. Dalam pengecekan kualitas dari model ini, digunakan metrik evaluasi *Accuracy*, *F1 Score*, *Recall Score*, *ROC AUC*, dan *Precision Score*. Skor yang didapatkan pada *training* dan *testing* ini sudah dirasa cukup, jadi tidak perlu dilakukan optimasi lagi.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

model.pkl

Gambar 3.21 Hasil Model *Customer Churn Prediction*

Setelah menganalisa hasil dari metrik evaluasi yang didapatkan, maka selanjutnya model *customer churn prediction* yang telah dibuat diekspor menjadi sebuah *file*. *File* model *customer churn prediction* di sini diberikan nama “model” dengan format *file* .pkl.

3.2.3 Realisasi Website Prediksi Churn Pelanggan

3.2.3.1 Fitur Website Prediksi Churn Pelanggan

A. Fitur Prediksi Data Tunggal

Di fitur ini, *user* akan diminta untuk memasukkan nilai-nilai dari setiap label yang ada pada laman *website*. Data yang *user input* merupakan data dari pelanggan yang ingin dilakukan prediksi churn.

Customer Churn Prediction

Area Name	Tangerang
Plan	Tanpa Plan
TV Plan	Tanpa Plan
Advance Promo	Tanpa Advance Promo
Complaint by Customer Service	0.00 / 200.00
Complaint by Email	0.00 / 250.00
Complaint by Social Media	0.00 / 200.00
Complaint by Telegram	0.00 / 100.00
Complaint by Whatsapp	0.00 / 400.00
Complaint by WIC	0.00 / 150.00
Predict	

Gambar 3.22 Tampilan Fitur Prediksi Data Tunggal



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

- 1) Melakukan *import library* yang dibutuhkan

```
import streamlit as st
import pandas as pd
# import numpy as np
import matplotlib
import pickle
import seaborn as sns
import six
import joblib
import sys
import os
import io
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
from google.cloud import storage
from matplotlib.backends.backend_pdf import PdfPages
from st_files_connection import FilesConnection
```

Gambar 3.23 Import Library

Langkah pertama yang harus dilakukan, yaitu *import* seluruh *library* yang diperlukan. Untuk membuat web prediksi ini dibutuhkan beberapa *library*, antara lain Streamlit, Pandas, Matplotlib, Pickle, Seaborn, Six, Joblib, Sys, Os, Io, Google Cloud Storage.

- 1) Menampilkan Selection Box
 - a. *Area Name*

```
user_area = st.selectbox('Area Name',
                        ('Tangerang', 'Depok', 'Cibubur', 'Bekasi', 'Surabaya',
                         'Semarang', 'Bogor', 'Malang', 'Palembang', 'Jakarta', 'Medan',
                         'Bandung', 'Bali', 'Makassar', 'Serang', 'Pekanbaru', 'Lampung',
                         'Solo', 'Cilegon', 'Karawang', 'Jambi'))
```

Gambar 3.24 Membuat Selection Box Area Name

Pada Selection Box untuk bagian *Area Name* terdapat 21 area. 21 area ini antara lain 'Tangerang', 'Depok', 'Cibubur', 'Bekasi', 'Surabaya', 'Semarang', 'Bogor', 'Malang', 'Palembang', 'Jakarta', 'Medan', 'Bandung', 'Bali', 'Makassar', 'Serang', 'Pekanbaru', 'Lampung', 'Solo', 'Cilegon', 'Karawang', 'Jambi'. Pembuatan Selection Box ini dilakukan dengan menggunakan *library* Streamlit, tepatnya dengan sintaks `st.selectbox()`. Select Box ini diinisialisasi dengan variabel bernama `user_area`.



b. *Plan*

```
user_plan = st.selectbox('Plan', ("jenis_plan"))
```

Gambar 3.25 Membuat Selection Box *Plan*

Pada Selection Box bagian *Plan*, terdapat 437 jenis plan. 437 jenis plan ini dapat dilihat pada L-1.

c. *Tv Plan*

```
tv_plan = st.selectbox('TV Plan', ("jenis_tv_plan"))
```

Gambar 3.26 Membuat Selection Box *Tv Plan*

Pada Selection Box bagian *Tv Plan*, terdapat 51 jenis plan. 51 jenis plan ini antara lain 'Tanpa *Plan*', 'Star*Tv*', 'Smar*Tv*+', 'Local Channel', 'STAR', 'Basic *Tv*', 'Cosmic *Tv*', 'Combo *Tv* (Fast50)', 'Smar*Tv*', 'No Channel', 'Combo *Tv* (Hype75)', 'SMART*Tv*', 'Combo *Tv* (Nova100)', 'Combo *Tv* (Value30)', 'Combo *Tv* (Gamer150)', 'Tv SOHO B (Movies and Sports)', 'Nova100 Mbps (Combo*Tv*+ELKMS)', 'SMART*Tv*+', 'Cosmic *Tv* SOHO', 'Advance *Tv*', 'Tv SOHO A (News and lifestyle)', 'Star *Tv* SOHO', 'Star *Tv* Plus', 'Star*Tv* Jet', 'Jet20 Mbps (Combo*Tv*)', 'Combo *Tv* (Jet20)', 'BASIC', 'Fast50 Mbps (Combo*Tv*+ELKMS)', 'Xtra Combo*Tv* (Fast50)', 'Hype75 Mbps (Combo*Tv*+ELKMS)', 'Xtra Combo*Tv* (Value30)', 'Xtra Combo*Tv* (Gamer150)', 'Value30 Mbps (Combo*Tv*+ELK)', 'Star *Tv* Plus (Android)', 'Xtra Combo*Tv* (Jet20)', 'Xtra Combo*Tv* (Nova100)', 'Xtra Combo*Tv* (Gamer75)', 'Xtra Combo*Tv* (Fast50 + ELKMS)', 'Xtra Combo*Tv* (V30 + ELK)', 'Xtra Combo*Tv* Jet20 (Android)', 'Combo *Tv* (Sonic150)', 'Combo*Tv* Pakubuwono', 'Combo*Tv* 77 Channel', 'Star *Tv*', 'No Channel (0)', 'Tv Only 60 Channel', 'Tv Starter 15+', 'Local Channel (Android)', 'Advance *Tv* (Android)', 'SOHO A', 'Combo*Tv* Pakubuwono (Android)'. Pembuatan Selection Box ini dilakukan dengan menggunakan *library* Streamlit, tepatnya dengan sintaks `st.selectbox()`. Select Box ini diinisialisasi dengan variabel bernama `tv_plan`.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

d. *Advance Promo*

```
adv_promo = st.selectbox('Advance Promo', ("jenis_advance_promo"))
```

Gambar 3.27 Membuat Selection Box *Advance Promo*

Pada Selection Box bagian *Advance Promo*, terdapat 378 jenis *Advance Promo*. 378 jenis *Advance Promo* ini dapat dilihat pada L-2.

2) Membuat *input* data numerikal dengan slider

a. Membuat slider untuk data *Complaint by Customer Service*

```
comp_cs = st.slider('Complaint by Customer Service', min_value = 0.0, max_value = 200.0 ,step=1.0)
```

Gambar 3.28 Membuat Slider *Complaint by Customer Service*

Dalam membuat *input* Slider menggunakan *library* Streamlit. Gambar 3.28 menunjukkan beberapa pengaturan, seperti angka minimum, angka maksimal, dan loncatan nilai dari setiap kali melakukan pergerakan terhadap Slider. Slider untuk *Complaint by Customer Service* ini diatur dengan nilai minimum 0, nilai maksimum 200, dan loncatan nilai 1.

b. Membuat slider untuk data *Complaint by Email*

```
comp_email = st.slider('Complaint by Email', min_value = 0.0, max_value = 250.0 ,step=1.0)
```

Gambar 3.29 Membuat Slider *Complaint by Email*

Dalam membuat *input* Slider menggunakan *library* Streamlit. Gambar 3.29 menunjukkan beberapa pengaturan, seperti angka minimum, angka maksimal, dan loncatan nilai dari setiap kali melakukan pergerakan terhadap Slider. Slider untuk *Complaint by Email* ini diatur dengan nilai minimum 0, nilai maksimum 250, dan loncatan nilai 1.

c. Membuat slider untuk data *Complaint by Social Media*

```
comp_socmed = st.slider('Complaint by Social Media', min_value = 0.0, max_value = 200.0 ,step=1.0)
```

Gambar 3.30 Membuat Slider *Complaint by Social Media*

Dalam membuat *input* Slider menggunakan *library* Streamlit. Gambar 3.30 menunjukkan beberapa pengaturan, seperti angka minimum, angka maksimal, dan loncatan nilai dari setiap kali melakukan pergerakan terhadap Slider. Slider untuk



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Complaint by Social Media ini diatur dengan nilai minimum 0, nilai maksimum 200, dan loncatan nilai 1.

- d. Membuat slider untuk data *Complaint by Telegram*

```
comp_tel = st.slider('Complaint by Telegram', min_value = 0.0, max_value = 100.0 ,step=1.0)
```

Gambar 3.31 Membuat Slider *Complaint by Telegram*

Dalam membuat *input* Slider menggunakan *library* Streamlit. Gambar 3.31 menunjukkan beberapa pengaturan, seperti angka minimum, angka maksimal, dan loncatan nilai dari setiap kali melakukan pergerakan terhadap Slider. Slider untuk *Complaint by Telegram* ini diatur dengan nilai minimum 0, nilai maksimum 100, dan loncatan nilai 1.

- e. Membuat slider untuk data *Complaint by Whatsapp*

```
comp_wa = st.slider('Complaint by Whatsapp', min_value = 0.0, max_value = 400.0 ,step=1.0)
```

Gambar 3.32 Membuat Slider *Complaint by Whatsapp*

Dalam membuat *input* Slider menggunakan *library* Streamlit. Gambar 3.32 menunjukkan beberapa pengaturan, seperti angka minimum, angka maksimal, dan loncatan nilai dari setiap kali melakukan pergerakan terhadap Slider. Slider untuk *Complaint by Telegram* ini diatur dengan nilai minimum 0, nilai maksimum 100, dan loncatan nilai 1.

- f. Membuat slider untuk data *Complaint by WIC*

```
comp_wic = st.slider('Complaint by WIC', min_value = 0.0, max_value = 150.0 ,step=1.0)
```

Gambar 3.33 Membuat Slider *Complaint by WIC*

Dalam membuat *input* Slider menggunakan *library* Streamlit. Gambar 3.33 menunjukkan beberapa pengaturan, seperti angka minimum, angka maksimal, dan loncatan nilai dari setiap kali melakukan pergerakan terhadap Slider. Slider untuk *Complaint by WIC* ini diatur dengan nilai minimum 0, nilai maksimum 150, dan loncatan nilai 1.

- 3) Membuat sebuah dataframe baru

Proses pembuatan dataframe baru ini ditujukan untuk memuat data yang *user input*.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
user_df_data = [[user_area,user_plan,tv_plan,adv_promo,comp_cs,comp_email,comp_socmed,comp_tel,
comp_wa,comp_wic]]
user_df_colnames = ["Area Name","Plan","Tv Plan","Advance Promo","Complaint by Customer Service",
"Complaint by Email","Complaint by Social Media","Complaint by Telegram","Complaint by Whatsapp",
"Complaint by WIC"]

input_df = pd.DataFrame(user_df_data,columns = user_df_colnames)
```

Gambar 3.34 Membuat Dataframe Baru

Pembuatan dataframe diawali dengan menyimpan seluruh data yang *user* masukkan pada sebuah variabel yang diberi nama *user_df_data*. Selanjutnya, perlu ditambahkan sebuah variabel yang menyimpan nama-nama dari kolom yang sesuai dengan *input* dari *user*. Setelah dua langkah tersebut dilakukan, selanjutnya proses penyatuan kedua data tersebut menjadi sebuah dataframe yang diberi nama dengan *input_df*. Rincian dari variabel *input_df*, yaitu *user_df_data* menjadi data dari dataframe *input_df* dan variabel *user_df_colnames* menjadi nama dari kolom – kolom dataframe *input_df*.

4) Membuat *dictionary* untuk proses *encoding data*

Pembuatan *dictionary* ini harus dilakukan karena pada proses prediksi churn, data yang diproses oleh model *machine learning* harus dalam bentuk numerical. Pembuatan *dictionary* ini dikhususkan untuk data berjenis kategorikal. Data kategorikal yang membutuhkan *dictionary* ini antara lain, *Area Name*, *Plan*, *Tv Plan*, *Advance Promo*, dan *Churn*. Khusus untuk data *Churn*, *dictionary* ini hanya diperlukan untuk proses *decoding* pada saat data ingin divisualisasikan.

a) *Dictionary Area Name*

```
area_dict = {'Bali': 0, 'Bandung': 1, 'Bekasi': 2, 'Bogor': 3, 'Cibubur': 4, 'Cilegon': 5, 'Cirebon': 6,
'Depok': 7, 'Jakarta': 8, 'Jambi': 9, 'Karawang': 10, 'Lampung': 11, 'Makassar': 12, 'Malang': 13,
'Medan': 14, 'Palembang': 15, 'Pekanbaru': 16, 'Purwokerto': 17, 'Semarang': 18, 'Serang': 19, 'Solo':
20, 'Surabaya': 21, 'Tangerang': 22, 'Tegal': 23}
```

Gambar 3.35 *Dictionary Area Name*

Dictionary yang dibuat untuk data *Area Name* ini diberi nama dengan *area_dict*. Di sini data pada *Area Name* dimasukkan dan diberikan kode sesuai dengan urutan *unique value* yang ada pada data *Area Name*.

b) *Dictionary Plan*

Dictionary yang dibuat untuk data *Plan* ini diberi nama dengan *area_dict*. Di sini data pada *Plan* dimasukkan dan diberikan kode sesuai dengan urutan *unique*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

value yang ada pada data *Plan*. *Dictionary Plan* dapat dilihat pada L-3. *Dictionary* yang dibuat untuk data *Plan* ini diberi nama dengan *plan_dict*. Di sini data pada *Plan* dimasukkan dan diberikan kode sesuai dengan urutan *unique value* yang ada pada data *Plan*.

c) *Dictionary Tv Plan*

Dictionary yang dibuat untuk data *Tv Plan* ini diberi nama dengan *area_dict*. Di sini data pada *Tv Plan* dimasukkan dan diberikan kode sesuai dengan urutan *unique value* yang ada pada data *Tv Plan*.

```
dict_tvplan = {'Advance TV': 0, 'Advance TV (Android)': 1, 'BASIC': 2, 'Basic TV': 3, 'Combo TV (Fast50)': 4, 'Combo TV (Gamer150)': 5, 'Combo TV (Hype75)': 6, 'Combo TV (Jet20)': 7, 'Combo TV (Nova100)': 8, 'Combo TV (Sonic150)': 9, 'Combo TV (Value30)': 10, 'ComboTV 77 Channel': 11, 'ComboTV Pakubuwono': 12, 'ComboTV Pakubuwono (Android)': 13, 'Cosmic TV': 14, 'Cosmic TV SOHO': 15, 'Fast50 Mbps (ComboTV+ELKMS)': 16, 'Hype75 Mbps (ComboTV+ELKMS)': 17, 'Jet20 Mbps (ComboTV)': 18, 'Local Channel': 19, 'Local Channel (Android)': 20, 'No Channel': 21, 'No Channel (0)': 22, 'Nova100 Mbps (ComboTV+ELKMS)': 23, 'SMARTV': 24, 'SMARTV+': 25, 'SOHO A': 26, 'STAR': 27, 'SmarTV': 28, 'SmarTV+': 29, 'Star TV': 30, 'Star TV Plus': 31, 'Star TV Plus (Android)': 32, 'Star TV SOHO': 33, 'StarTV': 34, 'StarTV Jet': 35, 'TV Only 60 Channel': 36, 'TV SOHO A (News and lifestyle)': 37, 'TV SOHO B (Movies and Sports)': 38, 'TV Starter 15+': 39, 'Tanpa Plan': 40, 'Value30 Mbps (ComboTV+ELK)': 41, 'Xtra ComboTV (Fast50 + ELKMS)': 42, 'Xtra ComboTV (Fast50)': 43, 'Xtra ComboTV (Gamer150)': 44, 'Xtra ComboTV (Gamer75)': 45, 'Xtra ComboTV (Jet20)': 46, 'Xtra ComboTV (Nova100)': 47, 'Xtra ComboTV (V30 + ELK)': 48, 'Xtra ComboTV (Value30)': 49, 'Xtra ComboTV Jet20 (Android)': 50}
```

Gambar 3.36 *Dictionary Tv Plan*

Dictionary yang dibuat untuk data *Tv Plan* ini diberi nama dengan *dict_tvplan*. Di sini data pada *Tv Plan* dimasukkan dan diberikan kode sesuai dengan urutan *unique value* yang ada pada data *Tv Plan*.

d) *Dictionary Advance Promo*

Dictionary yang dibuat untuk data *Advance Promo* ini diberi nama dengan *area_dict*. Di sini data pada *Advance Promo* dimasukkan dan diberikan kode sesuai dengan urutan *unique value* yang ada pada data *Advance Promo*. *Dictionary Advance Promo* dapat dilihat pada L-4. *Dictionary* yang dibuat untuk data *Advance Promo* ini diberi nama dengan *dict_promo*. Di sini data pada *Advance Promo* dimasukkan dan diberikan kode sesuai dengan urutan *unique value* yang ada pada data *Advance Promo*.

e) *Dictionary Churn*

```
dict_churn = {"Churn" : 1, "Not Churn" : 0}
```

Gambar 3.37 *Dictionary Churn*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Dictionary yang dibuat untuk data Churn ini diberi nama dengan `dict_churn`. Di sini data pada Churn dimasukkan dan diberikan kode sesuai dengan urutan *unique value* yang ada pada data Churn.

- 5) Membuat sebuah fungsi untuk *pre-processing data*

```
def preprocess_input(input_df):
    input_df['Area Name'] = input_df['Area Name'].map(area_dict)
    input_df['Plan'] = input_df['Plan'].map(plan_dict)
    input_df['Tv Plan'] = input_df['Tv Plan'].map(dict_tvplan)
    input_df['Advance Promo'] = input_df['Advance Promo'].map(dict_promo)
```

Gambar 3.38 Fungsi Pre-processing Data

Fungsi *pre-processing data* ini diberikan nama `preprocess_input`. Fungsi ini bekerja dengan memetakan ulang data – data berjenis kategorikal. Pemetaan ulang ini akan memasukkan data baru berupa hasil dari proses *encoding data* dan akan menimpa data lama dengan data hasil *encoding*. Kumpulan hasil data baru ini dimasukkan ke dalam suatu DataFrame baru bernama `input_df`.

- 6) Membuat sebuah fungsi untuk *decoding data*

```
def reverse(df):
    inverse_dict_areaname = {v: k for k, v in area_dict.items()}
    inverse_dict_plan = {v: k for k, v in plan_dict.items()}
    inverse_dict_tvplan = {v: k for k, v in dict_tvplan.items()}
    inverse_dict_promo = {v: k for k, v in dict_promo.items()}
    inverse_dict_churn = {v: k for k, v in dict_churn.items()}

    df['Area Name'] = df['Area Name'].map(inverse_dict_areaname)
    df['Plan'] = df['Plan'].map(inverse_dict_plan)
    df['Tv Plan'] = df['Tv Plan'].map(inverse_dict_tvplan)
    df['Advance Promo'] = df['Advance Promo'].map(inverse_dict_promo)
    df['Churn'] = df['Churn'].map(inverse_dict_churn)

    return(df)
```

Gambar 3.39 Fungsi Decoding Data

Cara kerja fungsi *decoding data* kurang lebih sama dengan fungsi *pre-processing data*. Di dalam fungsi ini dilakukan pembuatan *dictionary* baru yang memuat nilai berkebalikan dengan nilai *dictionary* awalnya. Setelah *dictionary* terbentuk, selanjutnya dilakukan pemetaan ulang sama yang cara kerjanya sama seperti fungsi *pre-processing data*.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

- 7) Membuat sebuah fungsi untuk prediksi

```
def predict_churn(data):
    data = preprocess_input(data)
    predictions = model.predict(data)
    data["Churn"] = predictions
    reverse_data = reverse(data)
    return data, reverse_data
```

Gambar 3.40 Fungsi Prediksi

Pembuatan fungsi ini nantinya akan berguna untuk prediksi data tunggal dan data *batch*. Dalam fungsi inilah model *machine learning* yang telah dibuat akan memproses data masuk yang diberikan oleh *user*. Hasil dari prediksi akan disimpan pada data dengan nama “Churn”.

- 8) Melakukan prediksi data tunggal

```
if st.button("Predict"):
    result_df = predict_churn(input_df)
    if result_df.iloc[0]['Churn'] == 0:
        st.success(f' The customer will be Not Churn')
    else:
        st.success(f'The customer will be Churn')
```

Gambar 3.41 Prediksi Data Tunggal

Proses diawali dengan membuat sebuah tombol dengan nama Predict. Tombol ini akan berfungsi sebagai eksekutor dari sebuah skema proses prediksi data. Diawali dengan pembuatan sebuah variabel yang berisikan eksekusi prediksi data *input_df* dengan fungsi *predict_churn* sebagai eksekutornya. Setelah hasil prediksi didapatkan, selanjutnya melihat hasil dengan bantuan *.iloc*. Jika keluaran berupa 0, maka *website* akan mengeksekusi sebuah perintah yang *outputnya* berupa teks bertulisan “The customer will be Not Churn”. Sebaliknya, jika keluaran berupa 1, maka *website* akan mengeksekusi sebuah perintah yang memiliki *output* berupa teks bertulisan “The customer will be Churn”.

B. Fitur Prediksi Data *Batch*

- 1) Membuat fungsi untuk *load data* churned



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Fungsi ini bekerja untuk menyeleksi data yang churned, lalu mengambil nilai-nilai tertentu yang nantinya akan dikeluarkan lagi dalam bentuk DataFrame dengan nama baru. Fungsi `load_non_churned` dapat dilihat di L-5.

- 2) Membuat fungsi untuk *load data* not churned

Fungsi ini bekerja untuk menyeleksi data yang churned, lalu mengambil nilai-nilai tertentu yang nantinya akan dikeluarkan lagi dalam bentuk DataFrame dengan nama baru. Fungsi `load_non_churned` dapat dilihat di L-6.

- 3) Membuat fungsi `dis_churn`

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan distribusi data churn dan tidak churn dari seluruh data hasil prediksi. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.

```
def dis_churn(data):
    pdf_pages1 = PdfPages("pie_chart_churn&notchurn.pdf")
    # st.header("Churn Distribution")
    fig, ax = plt.subplots()
    palette_color = sns.color_palette('rocket_r')

    churn_counts = data['Churn'].value_counts()
    ax = churn_counts.plot(kind='pie', autopct='%1.1f%%', colors=palette_color)
    ax.set_ylabel('')

    ax.legend(churn_counts.index)

    # Mengatur judul di tengah pie chart
    ax.set_title("Churn Distribution", loc='center')

    return fig, churn_counts
```

Gambar 3.42 Fungsi `dis_churn`

- 4) Membuat fungsi `area_proportion`

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan proporsi data *Area Name* yang mengalami churn. Grafik ini nantinya akan berisi 11 bagian, yaitu 10 data *Area Name* yang memiliki jumlah terbanyak dari data churn *Area Name* dan 1 data *Area Name* yang berisikan data *Area Name* yang tersisa. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def area_proportion(data):
    area_data_churned, plan_data_churned, tvplan_data_churned, \
    adv_data_churned, com_cs_data_churned, com_e_data_churned, \
    com_socmed_data_churned, tele_data_churned, wa_data_churned, wic_data_churned = load_churned(data)

    area_data_non_churned, plan_data_non_churned, tvplan_data_non_churned, \
    adv_data_non_churned, com_cs_data_non_churned, com_e_data_non_churned, \
    com_socmed_data_non_churned, tele_data_non_churned, wa_data_non_churned, wic_data_non_churned = \
    load_non_churned(data)

    area_data_merge = pd.merge(area_data_churned, area_data_non_churned, on="Area Name", how="outer")

    # Sort the DataFrame by 'Count Churned' in descending order
    area_data_merge.fillna(0, inplace=True)
    area_data_merge = area_data_merge.sort_values(by='Count Churned', ascending=False)

    # Select the top 10 rows
    top_10_churn = area_data_merge.head(10)

    # Calculate the sum of 'Count Churned' for the remaining rows
    remaining_sum_churn = area_data_merge.iloc[10:]['Count Churned'].sum()

    # Add the 'dll' row to the DataFrame
    dll_row_churn = {'Area Name': 'dll', 'Count Churned': remaining_sum_churn}
    top_10_churn = top_10_churn.append(dll_row_churn, ignore_index=True)

    # Plot the pie chart for churned
    plt.figure(figsize=(6, 6))
    patches, texts, autotexts = plt.pie(top_10_churn['Count Churned'], labels=None, autopct='%1.2f%%',
    startangle=140, colors=sns.color_palette("magma", len(top_10_churn)), textprops={'color': 'white'})
    plt.title('Proportion Count Churned by Area Name')
    plt.axis('equal')
    plt.legend(patches, top_10_churn['Area Name'], loc='center left', bbox_to_anchor=(0.5, -0.1), ncol=2)

    fig = plt.gcf()

    return fig, top_10_churn
```

Gambar 3.43 Fungsi area_proportion

5) Membuat fungsi area_proportion_1

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan proporsi data *Area Name* yang tidak mengalami churn. Grafik ini nantinya akan berisi 11 bagian, yaitu 10 data *Area Name* yang memiliki jumlah terbanyak dari data tidak churn *Area Name* dan 1 data *Area Name* yang berisikan data *Area Name* yang tersisa. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def area_proportion_1(data):
    area_data_churned, plan_data_churned, tvplan_data_churned, \
    adv_data_churned, com_cs_data_churned, com_e_data_churned, \
    com_socmed_data_churned, tele_data_churned, wa_data_churned, wic_data_churned = load_churned(data)

    area_data_non_churned, plan_data_non_churned, tvplan_data_non_churned, \
    adv_data_non_churned, com_cs_data_non_churned, com_e_data_non_churned, \
    com_socmed_data_non_churned, tele_data_non_churned, wa_data_non_churned, wic_data_non_churned = \
    load_non_churned(data)

    area_data_merge = pd.merge(area_data_churned, area_data_non_churned, on="Area Name", how="outer")
    # Sort the DataFrame by 'Count Not Churned' in descending order
    area_data_merge_notchurned = area_data_merge.sort_values(by='Count Not Churned', ascending=False)

    # Select the top 10 rows
    top_10_notchurn = area_data_merge_notchurned.head(10)

    # Calculate the sum of 'Count Not Churned' for the remaining rows
    remaining_sum_notchurn = area_data_merge_notchurned.iloc[10:]['Count Not Churned'].sum()

    # Add the 'dll' row to the DataFrame
    dll_row_notchurn = {'Area Name': 'dll', 'Count Not Churned': remaining_sum_notchurn}
    top_10_notchurn = top_10_notchurn.append(dll_row_notchurn, ignore_index=True)

    # Plot the pie chart for not churned
    plt.figure(figsize=(6, 6))
    patches, texts, autotexts = plt.pie(top_10_notchurn['Count Not Churned'], labels=None, autopct='%1.1f%', startangle=140, colors=sns.color_palette("magma", len(top_10_notchurn)), textprops={'color': 'white'})
    plt.title('Proportion Count Not Churned by Area Name')
    plt.axis('equal')
    plt.legend(patches, top_10_notchurn['Area Name'], loc='center left', bbox_to_anchor=(0.5, -0.1), ncol=2)

    fig = plt.gcf()

    return fig, top_10_notchurn
```

Gambar 3.44 Fungsi area_proportion_1

6) Membuat fungsi plan_proportion

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan proporsi data *Plan* yang mengalami churn. Grafik ini nantinya akan berisi 11 bagian, yaitu 10 data *Plan* yang memiliki jumlah terbanyak dari data churn *Plan* dan 1 data *Plan* yang berisikan data *Plan* yang tersisa. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def plan_proportion(data):
    area_data_churned, plan_data_churned, tvplan_data_churned, \
    adv_data_churned, com_cs_data_churned, com_e_data_churned, \
    com_socmed_data_churned, tele_data_churned, wa_data_churned, wic_data_churned = load_churned(data)

    area_data_non_churned, plan_data_non_churned, tvplan_data_non_churned, \
    adv_data_non_churned, com_cs_data_non_churned, com_e_data_non_churned, \
    com_socmed_data_non_churned, tele_data_non_churned, wa_data_non_churned, wic_data_non_churned = \
    load_non_churned(data)

    plan_data_merge = pd.merge(plan_data_churned, plan_data_non_churned, on="Plan", how="outer")

    # Sort the DataFrame by 'Count Churned' in descending order
    plan_data_merge.fillna(0, inplace=True)
    plan_data_merge = plan_data_merge.sort_values(by='Count Churned', ascending=False)

    # Select the top 10 rows
    top_10_churn = plan_data_merge.head(10)

    # Calculate the sum of 'Count Churned' for the remaining rows
    remaining_sum_churn = plan_data_merge.iloc[10:]['Count Churned'].sum()

    # Add the 'dll' row to the DataFrame
    dll_row_churn = {'Plan': 'dll', 'Count Churned': remaining_sum_churn}
    top_10_churn = top_10_churn.append(dll_row_churn, ignore_index=True)

    # Plot the pie chart for churned
    plt.figure(figsize=(6, 6))
    patches, texts, autotexts = plt.pie(top_10_churn['Count Churned'], labels=None, autopct='%1.2f%%',
    startangle=140, colors=sns.color_palette("magma", len(top_10_churn)), textprops={'color': 'white'})
    plt.title('Proportion Count Churned by Plan')
    plt.axis('equal')
    plt.legend(patches, top_10_churn['Plan'], loc='center left', bbox_to_anchor=(0.5, -0.1), ncol=2)

    fig = plt.gcf()

    return fig, top_10_churn
```

Gambar 3.45 Fungsi plan_proportion

7) Membuat fungsi plan_proportion_1

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan proporsi data *Plan* yang tidak mengalami churn. Grafik ini nantinya akan berisi 11 bagian, yaitu 10 data *Plan* yang memiliki jumlah terbanyak dari data tidak churn *Plan* dan 1 data *Plan* yang berisikan data *Plan* yang tersisa. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def plan_proportion_1(data):
    area_data_churned, plan_data_churned, tvplan_data_churned, \
    adv_data_churned, com_cs_data_churned, com_e_data_churned, \
    com_socmed_data_churned, tele_data_churned, wa_data_churned, wic_data_churned = load_churned(data)

    area_data_non_churned, plan_data_non_churned, tvplan_data_non_churned, \
    adv_data_non_churned, com_cs_data_non_churned, com_e_data_non_churned, \
    com_socmed_data_non_churned, tele_data_non_churned, wa_data_non_churned, wic_data_non_churned = \
    load_non_churned(data)

    plan_data_merge = pd.merge(plan_data_churned, plan_data_non_churned, on="Plan", how="outer")
    # Sort the DataFrame by 'Count Not Churned' in descending order
    plan_data_merge_notchurned = plan_data_merge.sort_values(by='Count Not Churned', ascending=False)

    # Select the top 10 rows
    top_10_notchurn = plan_data_merge_notchurned.head(10)

    # Calculate the sum of 'Count Not Churned' for the remaining rows
    remaining_sum_notchurn = plan_data_merge_notchurned.iloc[10:]['Count Not Churned'].sum()

    # Add the 'dll' row to the DataFrame
    dll_row_notchurn = {'Plan': 'dll', 'Count Not Churned': remaining_sum_notchurn}
    top_10_notchurn = top_10_notchurn.append(dll_row_notchurn, ignore_index=True)

    # Plot the pie chart for not churned
    plt.figure(figsize=(6, 6))
    patches, texts, autotexts = plt.pie(top_10_notchurn['Count Not Churned'], labels=None, autopct='%1.1f%%', startangle=140, colors=sns.color_palette("magma", len(top_10_notchurn)), textprops={'color': 'white'})
    plt.title('Proportion Count Not Churned by Plan')
    plt.axis('equal')
    plt.legend(patches, top_10_notchurn['Plan'], loc='center left', bbox_to_anchor=(0.5, -0.1), ncol=2)

    fig = plt.gcf()

    return fig, top_10_notchurn
```

Gambar 3.46 Fungsi plan_proportion_1

8) Membuat fungsi tvplan_proportion

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan proporsi data *Tv Plan* yang mengalami churn. Grafik ini nantinya akan berisi 11 bagian, yaitu 10 data *Tv Plan* yang memiliki jumlah terbanyak dari data churn *Tv Plan* dan 1 data *Tv Plan* yang berisikan data *Tv Plan* yang tersisa. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file PDF*.


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def tvplan_proportion(data):
    area_data_churned, plan_data_churned, tvplan_data_churned, \
    adv_data_churned, com_cs_data_churned, com_e_data_churned, \
    com_socmed_data_churned, tele_data_churned, wa_data_churned, wic_data_churned = load_churned(data)

    area_data_non_churned, plan_data_non_churned, tvplan_data_non_churned, \
    adv_data_non_churned, com_cs_data_non_churned, com_e_data_non_churned, \
    com_socmed_data_non_churned, tele_data_non_churned, wa_data_non_churned, wic_data_non_churned = \
    load_non_churned(data)

    tvplan_data_merge = pd.merge(tvplan_data_churned, tvplan_data_non_churned, on="Tv Plan", how="outer")

    # Sort the DataFrame by 'Count Churned' in descending order
    tvplan_data_merge = tvplan_data_merge.sort_values(by='Count Churned', ascending=False)

    # Select the top 10 rows
    top_10_churn = tvplan_data_merge.head(10)

    # Calculate the sum of 'Count Churned' for the remaining rows
    remaining_sum_churn = tvplan_data_merge.iloc[10:]['Count Churned'].sum()

    # Add the 'dll' row to the DataFrame
    dll_row_churn = {'Tv Plan': 'dll', 'Count Churned': remaining_sum_churn}
    top_10_churn = top_10_churn.append(dll_row_churn, ignore_index=True)

    # Plot the pie chart for Count Churned
    plt.figure(figsize=(6, 6))
    patches, texts, autotexts = plt.pie(top_10_churn['Count Churned'], labels=None, autopct='%1.1f%%',
    startangle=140, colors=sns.color_palette("magma", len(top_10_churn)), textprops={'color': 'white'})
    plt.title('Proportion Count Churned by Tv Plan')
    plt.axis('equal')
    plt.legend(patches, top_10_churn['Tv Plan'], loc='center left', bbox_to_anchor=(0.5, -0.1), ncol=2)

    # Display the churn pie chart using st.pyplot()
    fig = plt.gcf()

    return fig, top_10_churn
```

Gambar 3.47 Fungsi tvplan_proportion

9) Membuat fungsi tvplan_proportion_1

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan proporsi data *Tv Plan* yang tidak mengalami churn. Grafik ini nantinya akan berisi 11 bagian, yaitu 10 data *Tv Plan* yang memiliki jumlah terbanyak dari data tidak churn *Tv Plan* dan 1 data *Tv Plan* yang berisikan data *Tv Plan* yang tersisa. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def tvplan_proportion_1(data):
    area_data_churned, plan_data_churned, tvplan_data_churned, \
    adv_data_churned, com_cs_data_churned, com_e_data_churned, \
    com_socmed_data_churned, tele_data_churned, wa_data_churned, wic_data_churned = load_churned(data)

    area_data_non_churned, plan_data_non_churned, tvplan_data_non_churned, \
    adv_data_non_churned, com_cs_data_non_churned, com_e_data_non_churned, \
    com_socmed_data_non_churned, tele_data_non_churned, wa_data_non_churned, wic_data_non_churned = \
    load_non_churned(data)

    tvplan_data_merge = pd.merge(tvplan_data_churned, tvplan_data_non_churned, on="Tv Plan", how="outer")
    # Sort the DataFrame by 'Count Not Churned' in descending order
    tvplan_data_merge_notchurned = tvplan_data_merge.sort_values(by='Count Not Churned', ascending=False)

    # Select the top 10 rows
    top_10_notchurn = tvplan_data_merge_notchurned.head(10)

    # Calculate the sum of 'Count Not Churned' for the remaining rows
    remaining_sum_notchurn = tvplan_data_merge_notchurned.iloc[10:]['Count Churned'].sum()

    # Add the 'dll' row to the DataFrame
    dll_row_notchurn = {'Tv Plan': 'dll', 'Count Not Churned': remaining_sum_notchurn}
    top_10_notchurn = top_10_notchurn.append(dll_row_notchurn, ignore_index=True)

    # Plot the pie chart for Count Not Churned
    plt.figure(figsize=(6, 6))
    patches, texts, autotexts = plt.pie(top_10_notchurn['Count Not Churned'], labels=None, autopct='%1.1f%', startangle=140, colors=sns.color_palette("magma", len(top_10_notchurn)), textprops={'color': 'white'})
    plt.title('Proportion Count Not Churned by Tv Plan')
    plt.axis('equal')
    plt.legend(patches, top_10_notchurn['Tv Plan'], loc='center left', bbox_to_anchor=(0.5, -0.1), ncol=2)

    # Display the not churned pie chart using st.pyplot()
    fig = plt.gcf()
    return fig, top_10_notchurn
```

Gambar 3.48 Fungsi tvplan_proportion_1

10) Membuat fungsi advpro_proportion

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan proporsi data *Advance Promo* yang mengalami churn. Grafik ini nantinya akan berisi 11 bagian, yaitu 10 data *Advance Promo* yang memiliki jumlah terbanyak dari data churn *Advance Promo* dan 1 data *Advance Promo* yang berisikan data *Advance Promo* yang tersisa. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def advpro_proportion(data):
    area_data_churned, plan_data_churned, tvplan_data_churned, \
    adv_data_churned, com_cs_data_churned, com_e_data_churned, \
    com_socmed_data_churned, tele_data_churned, wa_data_churned, wic_data_churned = load_churned(data)

    area_data_non_churned, plan_data_non_churned, tvplan_data_non_churned, \
    adv_data_non_churned, com_cs_data_non_churned, com_e_data_non_churned, \
    com_socmed_data_non_churned, tele_data_non_churned, wa_data_non_churned, wic_data_non_churned = \
    load_non_churned(data)

    adv_data_merge = pd.merge(adv_data_churned, adv_data_non_churned, on="Advance Promo", how="outer")

    # Sort the DataFrame by 'Count Churned' in descending order
    adv_data_merge = adv_data_merge.sort_values(by='Count Churned', ascending=False)

    # Select the top 10 rows
    top_10_churn = adv_data_merge.head(10)

    # Calculate the sum of 'Count Churned' for the remaining rows
    remaining_sum_churn = adv_data_merge.iloc[10:]['Count Churned'].sum()

    # Add the 'dll' row to the DataFrame
    dll_row_churn = {'Advance Promo': 'dll', 'Count Churned': remaining_sum_churn}
    top_10_churn = top_10_churn.append(dll_row_churn, ignore_index=True)

    # Plot the pie chart for Count Churned
    plt.figure(figsize=(6, 6))
    patches, texts, autotexts = plt.pie(top_10_churn['Count Churned'], labels=None, autopct='%1.1f%%',
    startangle=140, colors=sns.color_palette("magma", len(top_10_churn)), textprops={'color': 'white'})
    plt.title('Proportion Count Churned by Advance Promo')
    plt.axis('equal')
    plt.legend(patches, top_10_churn['Advance Promo'], loc='center left', bbox_to_anchor=(0.5, -0.1),
    ncol=2)

    fig = plt.gcf()

    return fig, top_10_churn
```

Gambar 3.49 Fungsi advpro_proportion

11) Membuat fungsi advpro_proportion_1

Fungsi ini bekerja untuk mengatur grafik pie yang berisikan proporsi data *Advance Promo* yang tidak mengalami churn. Grafik ini nantinya akan berisi 11 bagian, yaitu 10 data *Advance Promo* yang memiliki jumlah terbanyak dari data tidak churn *Advance Promo* dan 1 data *Advance Promo* yang berisikan data *Advance Promo* yang tersisa. Grafik pie ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def advpro_proportion_1(data):
    area_data_churned, plan_data_churned, tvplan_data_churned, \
    adv_data_churned, com_cs_data_churned, com_e_data_churned, \
    com_socmed_data_churned, tele_data_churned, wa_data_churned, wic_data_churned = load_churned(data)

    area_data_non_churned, plan_data_non_churned, tvplan_data_non_churned, \
    adv_data_non_churned, com_cs_data_non_churned, com_e_data_non_churned, \
    com_socmed_data_non_churned, tele_data_non_churned, wa_data_non_churned, wic_data_non_churned = \
    load_non_churned(data)

    adv_data_merge = pd.merge(adv_data_churned, adv_data_non_churned, on="Advance Promo", how="outer")
    # Sort the DataFrame by 'Count Not Churned' in descending order
    adv_data_merge_notchurned = adv_data_merge.sort_values(by='Count Not Churned', ascending=False)

    # Select the top 10 rows
    top_10_notchurn = adv_data_merge_notchurned.head(10)

    # Calculate the sum of 'Count Not Churned' for the remaining rows
    remaining_sum_notchurn = adv_data_merge_notchurned.iloc[10:]['Count Churned'].sum()

    # Add the 'dll' row to the DataFrame
    dll_row_notchurn = {'Advance Promo': 'dll', 'Count Not Churned': remaining_sum_notchurn}
    top_10_notchurn = top_10_notchurn.append(dll_row_notchurn, ignore_index=True)

    # Plot the pie chart for Count Not Churned
    plt.figure(figsize=(6, 6))
    patches, texts, autotexts = plt.pie(top_10_notchurn['Count Not Churned'], labels=None, autopct='%1.1f%',
    startangle=140, colors=sns.color_palette("magma", len(top_10_notchurn)), textprops={'color':
    'white'})
    plt.title('Proportion Count Not Churned by Advance Promo')
    plt.axis('equal')
    plt.legend(patches, top_10_notchurn['Advance Promo'], loc='center left', bbox_to_anchor=(0.5, -0.1),
    ncol=2)

    fig = plt.gcf()
    return fig, top_10_notchurn
```

Gambar 3.50 Fungsi advpro_proportion_1

12) Membuat fungsi full_area

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Area Name* secara menyeluruh. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def full_area(data):
    unique_area_name = sorted(data['Area Name'].unique())
    area_name_counts = data['Area Name'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Area Name'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Area Name'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Area Name': unique_area_name,
                              'Churn': [churn_counts.get(area, 0) for area in unique_area_name],
                              'Not Churn': [not_churn_counts.get(area, 0) for area in unique_area_name]})

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Area Name', y='value', hue='variable', data=pd.melt(churn_data, ['Area Name']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Area')
    plt.xlabel('Area Name')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()

    return fig
```

Gambar 3.51 Fungsi full_area

13) Membuat fungsi five_area

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Area Name* yang sudah dilakukan proses *ranking*. Proses *ranking* ini akan mengurutkan data “Total Churn + Total Not Churn” terbanyak hingga terkecil dan mengambilnya 5 data teratas. Data “Total Churn + Total Not Churn” diperoleh dari proses penjumlahan data “Total Churn” dan data “Total Not Churn”. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def five_area(data):
    unique_area_name = sorted(data['Area Name'].unique())
    area_name_counts = data['Area Name'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Area Name'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Area Name'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Area Name': unique_area_name,
                              'Churn': [churn_counts.get(area, 0) for area in unique_area_name],
                              'Not Churn': [not_churn_counts.get(area, 0) for area in unique_area_name]})

    # Combine 'Churn' and 'Not Churn' counts to get the total churn + not churn counts
    churn_data['Total Churn + Not Churn'] = churn_data['Churn'] + churn_data['Not Churn']

    # Sort the DataFrame based on the 'Total Churn + Not Churn' column in descending order
    sorted_churn_data = churn_data.sort_values(by='Total Churn + Not Churn', ascending=False)

    # # Get the top 5 areas with the highest total churn + not churn counts

    top_5_areas = sorted_churn_data.head(5)
    top_5_areas = top_5_areas.drop(columns=['Total Churn + Not Churn'])

    churn_data = top_5_areas # Hapus kolom dari churn_data juga

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Area Name', y='value', hue='variable', data=pd.melt(churn_data, ['Area Name']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Area')
    plt.xlabel('Area Name')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.52 Fungsi five_area

14) Membuat fungsi ten_area

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Area Name* yang sudah dilakukan proses *ranking*. Proses *ranking* ini akan mengurutkan data “Total Churn + Total Not Churn” terbanyak hingga terkecil dan mengambilnya 10 data teratas. Data “Total Churn + Total Not Churn” diperoleh dari proses penjumlahan data “Total Churn” dan data “Total Not Churn”. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file PDF*.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def ten_area(data):
    unique_area_name = sorted(data['Area Name'].unique())
    area_name_counts = data['Area Name'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Area Name'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Area Name'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Area Name': unique_area_name,
                              'Churn': [churn_counts.get(area, 0) for area in unique_area_name],
                              'Not Churn': [not_churn_counts.get(area, 0) for area in unique_area_name]})

    # Combine 'Churn' and 'Not Churn' counts to get the total churn + not churn counts
    churn_data['Total Churn + Not Churn'] = churn_data['Churn'] + churn_data['Not Churn']

    # Sort the DataFrame based on the 'Total Churn + Not Churn' column in descending order
    sorted_churn_data = churn_data.sort_values(by='Total Churn + Not Churn', ascending=False)

    # Get the top 5 areas with the highest total churn + not churn counts
    top_10_areas = sorted_churn_data.head(10)
    top_10_areas = top_10_areas.drop(columns=['Total Churn + Not Churn'])

    churn_data = top_10_areas
    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Area Name', y='value', hue='variable', data=pd.melt(churn_data, ['Area Name']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Area')
    plt.xlabel('Area Name')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.53 Fungsi ten_area

15) Membuat fungsi full_plan

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Plan* secara menyeluruh. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def full_plan(data):
    unique_plan = sorted(data['Plan'].unique())
    plan_counts = data['Plan'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Plan'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Plan'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Plan': unique_plan,
                              'Churn': [churn_counts.get(plan, 0) for plan in unique_plan],
                              'Not Churn': [not_churn_counts.get(plan, 0) for plan in
                                             unique_plan]})

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Plan', y='value', hue='variable', data=pd.melt(churn_data, ['Plan']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Plan')
    plt.xlabel('Plan')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.54 Fungsi full_plan

16) Membuat fungsi five_plan

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Plan* yang sudah dilakukan proses *ranking*. Proses *ranking* ini akan mengurutkan data “Total Churn + Total Not Churn” terbanyak hingga terkecil dan mengambilnya 5 data teratas. Data “Total Churn + Total Not Churn” diperoleh dari proses penjumlahan data “Total Churn” dan data “Total Not Churn”. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def five_plan(data):
    unique_plan = sorted(data['Plan'].unique())
    plan_counts = data['Plan'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Plan'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Plan'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Plan': unique_plan,
                              'Churn': [churn_counts.get(plan, 0) for plan in unique_plan],
                              'Not Churn': [not_churn_counts.get(plan, 0) for plan in
                                             unique_plan]})

    # Combine 'Churn' and 'Not Churn' counts to get the total churn + not churn counts
    churn_data['Total Churn + Not Churn'] = churn_data['Churn'] + churn_data['Not Churn']

    # Sort the DataFrame based on the 'Total Churn + Not Churn' column in descending order
    sorted_churn_data = churn_data.sort_values(by='Total Churn + Not Churn', ascending=False)

    # Get the top 5 areas with the highest total churn + not churn counts
    top_5_areas = sorted_churn_data.head(5)
    top_5_areas = top_5_areas.drop(columns=['Total Churn + Not Churn'])

    churn_data = top_5_areas

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Plan', y='value', hue='variable', data=pd.melt(churn_data, ['Plan']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Plan')
    plt.xlabel('Plan')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.55 Fungsi five_plan

17) Membuat fungsi ten_plan

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Plan* yang sudah dilakukan proses *ranking*. Proses *ranking* ini akan mengurutkan data “Total Churn + Total Not Churn” terbanyak hingga terkecil dan mengambilnya 10 data teratas. Data “Total Churn + Total Not Churn” diperoleh dari proses penjumlahan data “Total Churn” dan data “Total Not Churn”. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def ten_plan(data):
    unique_plan = sorted(data['Plan'].unique())
    plan_counts = data['Plan'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Plan'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Plan'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Plan': unique_plan,
                              'Churn': [churn_counts.get(plan, 0) for plan in unique_plan],
                              'Not Churn': [not_churn_counts.get(plan, 0) for plan in
                                             unique_plan]})

    # Combine 'Churn' and 'Not Churn' counts to get the total churn + not churn counts
    churn_data['Total Churn + Not Churn'] = churn_data['Churn'] + churn_data['Not Churn']

    # Sort the DataFrame based on the 'Total Churn + Not Churn' column in descending order
    sorted_churn_data = churn_data.sort_values(by='Total Churn + Not Churn', ascending=False)

    # Get the top 5 areas with the highest total churn + not churn counts
    top_10_areas = sorted_churn_data.head(10)
    top_10_areas = top_10_areas.drop(columns=['Total Churn + Not Churn'])

    churn_data = top_10_areas

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Plan', y='value', hue='variable', data=pd.melt(churn_data, ['Plan']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Plan')
    plt.xlabel('Plan')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.56 Fungsi ten_plan

18) Membuat fungsi full_tvplan

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Tv Plan* secara menyeluruh. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def full_tvplan(data):
    unique_tvplan = sorted(data['Tv Plan'].unique())
    tvplan_counts = data['Tv Plan'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Tv Plan'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Tv Plan'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Tv Plan': unique_tvplan,
                              'Churn': [churn_counts.get(tvplan, 0) for tvplan in
                                         unique_tvplan],
                              'Not Churn': [not_churn_counts.get(tvplan, 0) for tvplan in
                                             unique_tvplan]})

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Tv Plan', y='value', hue='variable', data=pd.melt(churn_data, ['Tv Plan']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Tv Plan')
    plt.xlabel('Tv Plan')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.57 Fungsi full_tvplan

19) Membuat fungsi five_tvplan

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Tv Plan* yang sudah dilakukan proses *ranking*. Proses *ranking* ini akan mengurutkan data “Total Churn + Total Not Churn” terbanyak hingga terkecil dan mengambilnya 5 data teratas. Data “Total Churn + Total Not Churn” diperoleh dari proses penjumlahan data “Total Churn” dan data “Total Not Churn”. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def five_tvplan(data):
    unique_tvplan = sorted(data['Tv Plan'].unique())
    tvplan_counts = data['Tv Plan'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Tv Plan'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Tv Plan'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Tv Plan': unique_tvplan,
                              'Churn': [churn_counts.get(plan, 0) for plan in unique_tvplan],
                              'Not Churn': [not_churn_counts.get(plan, 0) for plan in
                                             unique_tvplan]})

    # Combine 'Churn' and 'Not Churn' counts to get the total churn + not churn counts
    churn_data['Total Churn + Not Churn'] = churn_data['Churn'] + churn_data['Not Churn']

    # Sort the DataFrame based on the 'Total Churn + Not Churn' column in descending order
    sorted_churn_data = churn_data.sort_values(by='Total Churn + Not Churn', ascending=False)

    # Get the top 5 areas with the highest total churn + not churn counts
    top_5_areas = sorted_churn_data.head(5)
    top_5_areas = top_5_areas.drop(columns=['Total Churn + Not Churn'])

    churn_data = top_5_areas

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Tv Plan', y='value', hue='variable', data=pd.melt(churn_data, ['Tv Plan']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Tv Plan')
    plt.xlabel('Tv Plan')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.58 Fungsi five_tvplan

20) Membuat fungsi ten_tvplan

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Tv Plan* yang sudah dilakukan proses *ranking*. Proses *ranking* ini akan mengurutkan data “Total Churn + Total Not Churn” terbanyak hingga terkecil dan mengambilnya 10 data teratas. Data “Total Churn + Total Not Churn” diperoleh dari proses penjumlahan data “Total Churn” dan data “Total Not Churn”. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def ten_tvplan(data):
    unique_tvplan = sorted(data['Tv Plan'].unique())
    tvplan_counts = data['Tv Plan'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Tv Plan'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Tv Plan'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Tv Plan': unique_tvplan,
                              'Churn': [churn_counts.get(plan, 0) for plan in unique_tvplan],
                              'Not Churn': [not_churn_counts.get(plan, 0) for plan in
                                             unique_tvplan]})

    # Combine 'Churn' and 'Not Churn' counts to get the total churn + not churn counts
    churn_data['Total Churn + Not Churn'] = churn_data['Churn'] + churn_data['Not Churn']

    # Sort the DataFrame based on the 'Total Churn + Not Churn' column in descending order
    sorted_churn_data = churn_data.sort_values(by='Total Churn + Not Churn', ascending=False)

    # Get the top 5 areas with the highest total churn + not churn counts
    top_10_areas = sorted_churn_data.head(10)
    top_10_areas = top_10_areas.drop(columns=['Total Churn + Not Churn'])

    churn_data = top_10_areas

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Tv Plan', y='value', hue='variable', data=pd.melt(churn_data, ['Tv Plan']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Tv Plan')
    plt.xlabel('Tv Plan')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.59 Fungsi ten_tvplan

21) Membuat fungsi full_adv

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Advance Promo* secara menyeluruh. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file PDF*.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def full_adv(data):
    unique_adv_promo = sorted(data['Advance Promo'].unique())
    adv_promo_counts = data['Advance Promo'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Advance Promo'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Advance Promo'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Advance Promo': unique_adv_promo,
                              'Churn': [churn_counts.get(adv_promo, 0) for adv_promo in
                                         unique_adv_promo],
                              'Not Churn': [not_churn_counts.get(adv_promo, 0) for adv_promo in
                                             unique_adv_promo]})

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Advance Promo', y='value', hue='variable', data=pd.melt(churn_data, ['Advance
    Promo']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Advance Promo')
    plt.xlabel('Advance Promo')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.60 Fungsi full_adv

22) Membuat fungsi five_adv

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Advance Promo* yang sudah dilakukan proses *ranking*. Proses *ranking* ini akan mengurutkan data “Total Churn + Total Not Churn” terbanyak hingga terkecil dan mengambilnya 5 data teratas. Data “Total Churn + Total Not Churn” diperoleh dari proses penjumlahan data “Total Churn” dan data “Total Not Churn”. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def five_adv(data):
    unique_adv_promo = sorted(data['Advance Promo'].unique())
    adv_promo_counts = data['Advance Promo'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Advance Promo'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Advance Promo'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Advance Promo': unique_adv_promo,
                              'Churn': [churn_counts.get(adv_promo, 0) for adv_promo in
                                         unique_adv_promo],
                              'Not Churn': [not_churn_counts.get(adv_promo, 0) for adv_promo in
                                             unique_adv_promo]})

    # Combine 'Churn' and 'Not Churn' counts to get the total churn + not churn counts
    churn_data['Total Churn + Not Churn'] = churn_data['Churn'] + churn_data['Not Churn']

    # Sort the DataFrame based on the 'Total Churn + Not Churn' column in descending order
    sorted_churn_data = churn_data.sort_values(by='Total Churn + Not Churn', ascending=False)

    # Get the top 5 areas with the highest total churn + not churn counts
    top_5_areas = sorted_churn_data.head(5)
    top_5_areas = top_5_areas.drop(columns=['Total Churn + Not Churn'])

    churn_data = top_5_areas

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Advance Promo', y='value', hue='variable', data=pd.melt(churn_data, ['Advance
    Promo']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Advance Promo')
    plt.xlabel('Advance Promo')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.61 Fungsi five_adv

23) Membuat fungsi ten_adv

Fungsi ini bekerja untuk mengatur grafik bar yang berisikan proporsi churn dan tidak churn dari data *Advance Promo* yang sudah dilakukan proses *ranking*. Proses *ranking* ini akan mengurutkan data “Total Churn + Total Not Churn” terbanyak hingga terkecil dan mengambilnya 10 data teratas. Data “Total Churn + Total Not Churn” diperoleh dari proses penjumlahan data “Total Churn” dan data “Total Not Churn”. Grafik ini dapat digunakan dalam proses menampilkan grafik tersebut dan dapat dibuat menjadi *file* PDF.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def ten_adv(data):
    unique_adv_promo = sorted(data['Advance Promo'].unique())
    adv_promo_counts = data['Advance Promo'].value_counts()

    # Hitung jumlah Churn dan Not Churn untuk setiap area
    churn_counts = data[data['Churn'] == 'Churn']['Advance Promo'].value_counts()
    not_churn_counts = data[data['Churn'] == 'Not Churn']['Advance Promo'].value_counts()

    # Membuat dataframe untuk menyimpan hasil perhitungan
    churn_data = pd.DataFrame({'Advance Promo': unique_adv_promo,
                              'Churn': [churn_counts.get(adv_promo, 0) for adv_promo in
                                         unique_adv_promo],
                              'Not Churn': [not_churn_counts.get(adv_promo, 0) for adv_promo in
                                             unique_adv_promo]})

    # Combine 'Churn' and 'Not Churn' counts to get the total churn + not churn counts
    churn_data['Total Churn + Not Churn'] = churn_data['Churn'] + churn_data['Not Churn']

    # Sort the DataFrame based on the 'Total Churn + Not Churn' column in descending order
    sorted_churn_data = churn_data.sort_values(by='Total Churn + Not Churn', ascending=False)

    # Get the top 5 areas with the highest total churn + not churn counts
    top_10_areas = sorted_churn_data.head(10)
    top_10_areas = top_10_areas.drop(columns=['Total Churn + Not Churn'])

    churn_data = top_10_areas

    # Membuat plot menggunakan sns.catplot
    sns.set(style="whitegrid")
    plt.figure(figsize=(12, 6))
    plot = sns.catplot(x='Advance Promo', y='value', hue='variable', data=pd.melt(churn_data, ['Advance
    Promo']),
                      kind='bar', height=6, aspect=2.5, palette='magma')
    plt.title('Proporsi Churn dan Not Churn berdasarkan Advance Promo')
    plt.xlabel('Advance Promo')
    plt.ylabel('Jumlah')
    plt.xticks(rotation=90)

    fig = plt.gcf()
    return fig
```

Gambar 3.62 Fungsi ten_adv

24) Membuat fungsi combine_pdfs

Fungsi ini bekerja untuk mengatur seluruh grafik yang telah dibuat dan membuatnya menjadi *file* PDF. Semua *file* PDF yang telah dibuat akan digabungkan menjadi satu. Nantinya *file* gabungan ini akan dapat diunduh. Fungsi *comine_pdfs* dapat dilihat pada L-7.

25) Membuat fungsi download_combined_pdf

Fungsi ini berkerja untuk mengatur sebuah tombol yang nantinya akan memunculkan tombol lain. Tombol yang muncul ini nantinya akan mengunduh *file* hasil gabungan yang ada di fungsi *combine_pdfs*.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
def download_combined_pdf(data):
    # Panggil fungsi untuk melakukan kombinasi PDF
    combine_pdfs(data)

    if st.button("Click Here"):
        with open("combined_report.pdf", "rb") as file:
            pdf_bytes = file.read()
            st.download_button("Download Churn Report", pdf_bytes, file_name="churn_report.pdf",
                               key="pdf_download")
```

Gambar 3.63 Fungsi download_combined_pdf

26) Membuat fungsi visualized_data_batch

Fungsi ini akan menampilkan tampilan dari halaman "Batch Prediction". Di dalam halaman ini akan terdapat visualisasi data dari seluruh data yang telah diprediksi. Visualisasi data ini berupa grafik pie dan grafik batang. Pada halaman ini juga memvisualkan data hasil prediksi dalam bentuk tabel. Tabel ini nantinya dapat diunduh. File yang dihasilkan akan berformat .csv. File CSV ini nantinya akan dapat diproses lebih lanjut oleh divisi terkait, sesuai dengan kebutuhannya. Fungsi visualized_data_batch dapat dilihat pada L-8.

3.2.3.2 Cloud dan Hosting Website

A. *Connect* Google Cloud

- 1) Masuk ke console.cloud.google.com dan lakukan login
- 2) Masukkan email dan password akun Google



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Gambar 3.64 Login Akun Google – 1

Pada laman ini, isikan email atau nomor telepon sesuai dengan akun yang terdaftar. Selanjutnya, pada laman ini, sesuai dengan Gambar 3.65 ini, isikan password sesuai dengan akun yang terdaftar.



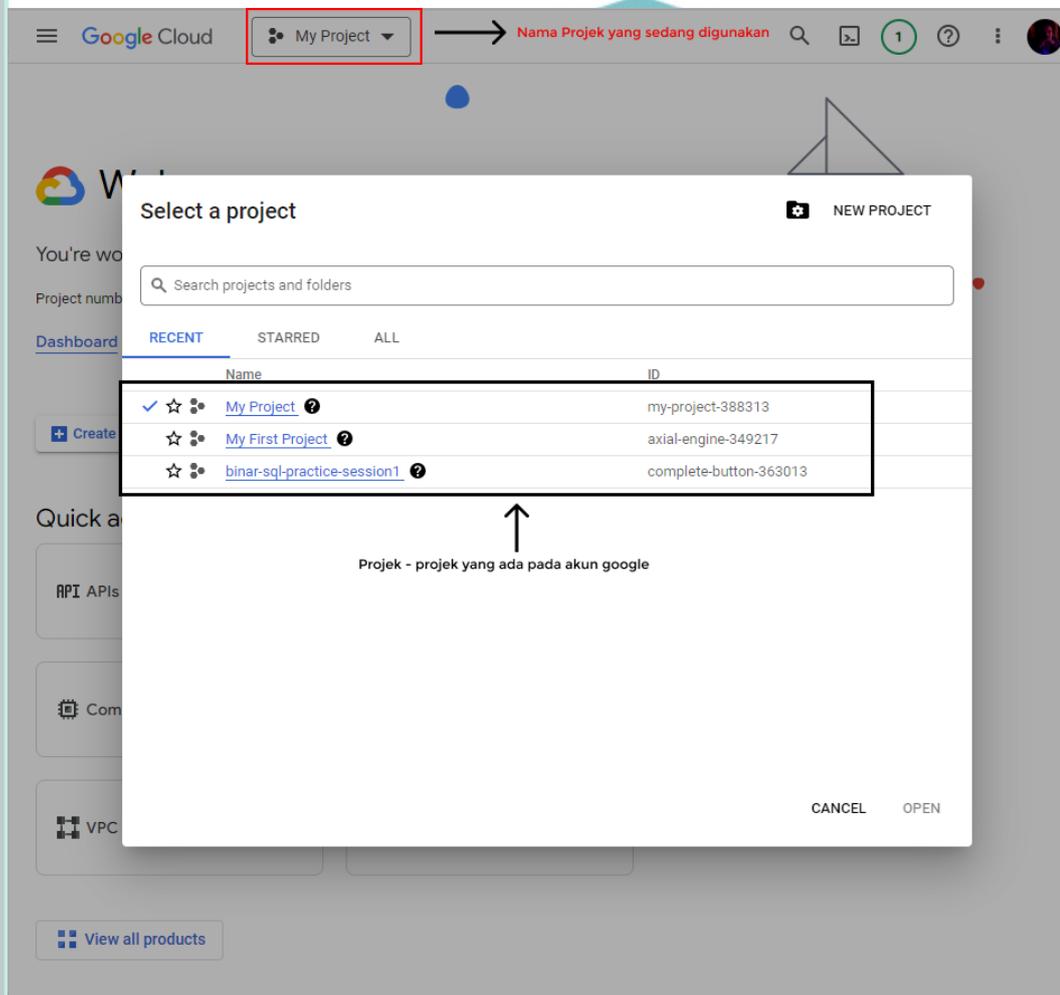
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Gambar 3.65 Login Akun Google – 2

- 3) Masuk ke proyek yang sudah disiapkan

Jika sudah melakukan login, masuk ke proyek yang ada pada Console Google Cloud Platform. Berikut adalah tampilannya.



Gambar 3.66 Pilih Projek

- 4) Pilihlah “Buckets” pada “Cloud Storage” Service

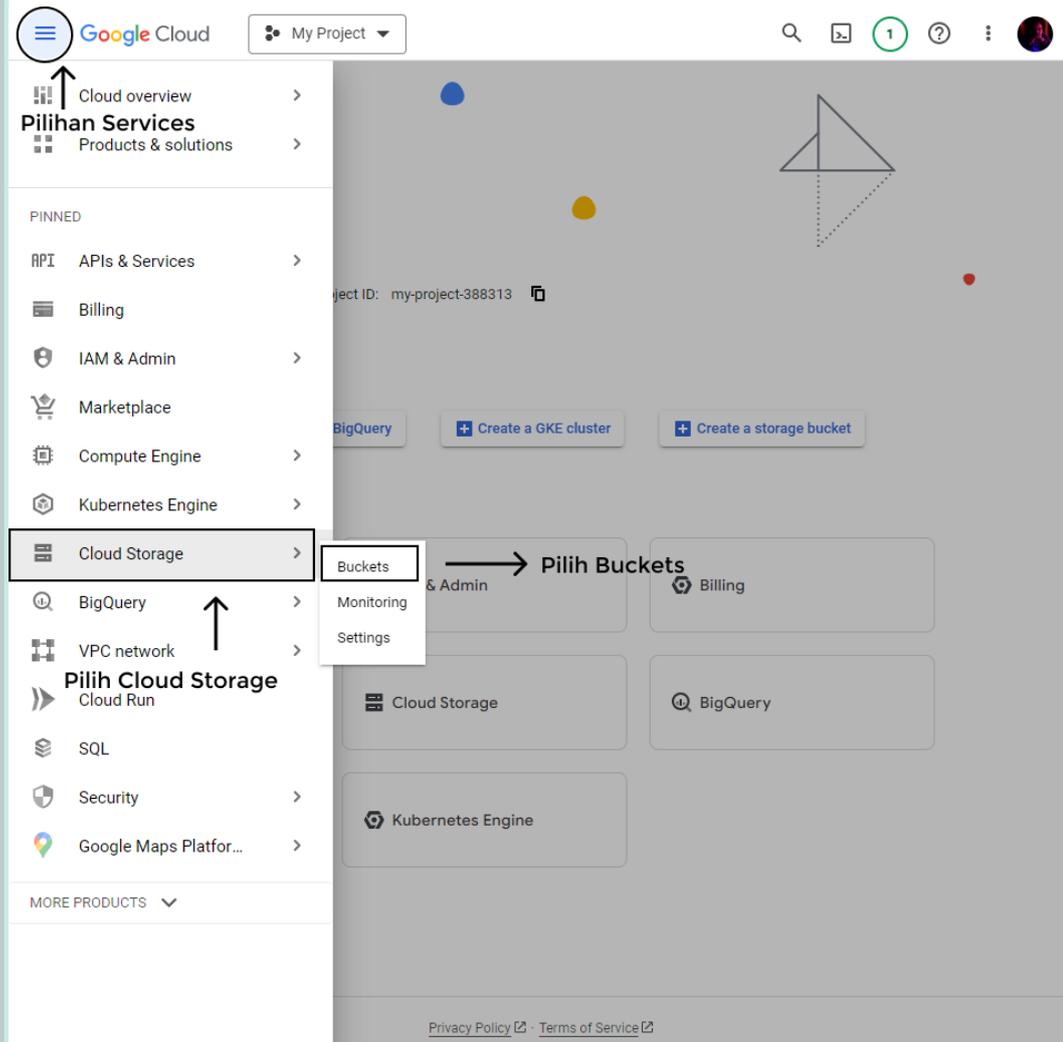
Setelah memilih proyek yang sesuai, selanjutnya masuklah ke halaman Bucket pada GCP. Gambar 3.67 ini menunjukkan cara untuk menuju halaman Bucket.



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 3.67 Menuju Halaman Bucket

5) Membuat Bucket

Setelah berada di halaman Bucket, langkah selanjutnya adalah membuat Bucket. Pada laman Bucket, terdapat pilihan untuk membuat Bucket baru. Klik tombol CREATE untuk membuat Bucket baru. Gambar berikut menunjukkan tombol untuk membuat Bucket.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Klik tombol CREATE

Google Cloud My Project

Cloud Storage Buckets **CREATE** REFRESH HELP ASSISTANT LEARN

Try The New Cloud Storage Monitoring Dashboard

Check out the new Cloud Storage monitoring dashboard and Bucket Observability pages! Powered by Cloud Operations, you can customize these dashboards for each project.

[TRY NOW](#)

View security recommendations

Improve security by applying security recommendations to your buckets. The security insights column in the table describes which buckets have excess permissions.

[VIEW IN TABLE](#) [LEARN MORE](#)

Filter Filter buckets

<input type="checkbox"/>	Name ↑	Created	Location type	Loca
<input type="checkbox"/>	artifacts.my-project-388313.appspot.com	Jul 12, 2023, 12:07:52 AM	Multi-region	us
<input type="checkbox"/>	model_churn	Jul 12, 2023, 1:31:33 AM	Region	asia-
<input type="checkbox"/>	my-project-388313.cloudbuild	Jul 12, 2023, 12:01:27 AM	Multi-region	us
<input type="checkbox"/>	parkingsystem2023	Jun 7, 2023, 11:09:31 AM	Region	asia-

Gambar 3.68 Tombol Menambahkan Bucket

Selanjutnya akan tampil sebuah halaman untuk mengatur Bucket. Isikan nama Bucket sesuai dengan keinginan dan pilihlah Location Type dengan Region asia-southeast2 (Jakarta).

**POLITEKNIK
NEGERI
JAKARTA**



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

← Create a bucket

Name your bucket

Pick a globally unique, permanent name. [Naming guidelines](#)

ml_model_skripsi

Tip: Don't include any sensitive information. Isikan nama model

▼ LABELS (OPTIONAL)

[CONTINUE](#)

Choose where to store your data

This choice defines the geographic placement of your data and affects cost, performance, and availability. Cannot be changed later. [Learn more](#)

Location type

Multi-region
Highest availability across largest area

Dual-region
High availability and low latency across 2 regions

Region
Lowest latency within a single region

asia-northeast2 (Osaka)

asia-northeast3 (Seoul)

asia-south1 (Mumbai)

asia-south2 (Delhi)

asia-southeast1 (Singapore)

asia-southeast2 (Jakarta) Pilih Region sesuai dengan wilayah, Jakarta

australia-southeast1 (Sydney)

australia-southeast2 (Melbourne)

Good to know

Location pricing

Storage rates vary depending on the storage class of your data and location of your bucket. [Pricing details](#)

Current configuration: Region / Standard

Item	Cost
us-east1 (South Carolina)	\$0.020 per GB-month

ESTIMATE YOUR MONTHLY COST

Gambar 3.69 Panduan Mengatur Bucket – 1

Jika semua hal sudah diatur sesuai dengan kebutuhan, selanjutnya tekan tombol CREATE untuk membuat Bucket ini.

POLITEKNIK
NEGERI
JAKARTA



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

• Choose where to store your data

This choice defines the geographic placement of your data and affects cost, performance, and availability. Cannot be changed later. [Learn more](#)

Location type

- Multi-region
Highest availability across largest area
- Dual-region
High availability and low latency across 2 regions
- Region
Lowest latency within a single region

asia-southeast2 (Jakarta)

CONTINUE

• Choose a storage class for your data

Default storage class: Standard

• Choose how to control access to objects

Public access prevention: On

Access control: Uniform

• Choose how to protect object data

Protection tools: None

Data encryption: Google-managed

Buat Bucket dengan klik tombol CREATE

CREATE

CANCEL

Gambar 3.70 Panduan Mengatur Bucket – 2

6) Melakukan *upload file* model

← Bucket details

ml_model_skripsi Nama Bucket yang sudah dibuat

Location	Storage class	Public access	Protection
asia-southeast2 (Jakarta)	Standard	Not public	None

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY INVENTORY REPORTS

Buckets > ml_model_skripsi

Upload file model dengan klik tombol **UPLOAD FILES**

UPLOAD FILES **UPLOAD FOLDER** **CREATE FOLDER** **TRANSFER DATA** **MANAGE HOLDS** **DOWNLOAD** **DELETE**

Filter by name prefix only **Filter** Filter objects and folders

Name	Size	Type	Created	Storage class	Last modified	Public access	Version history
No rows to display							

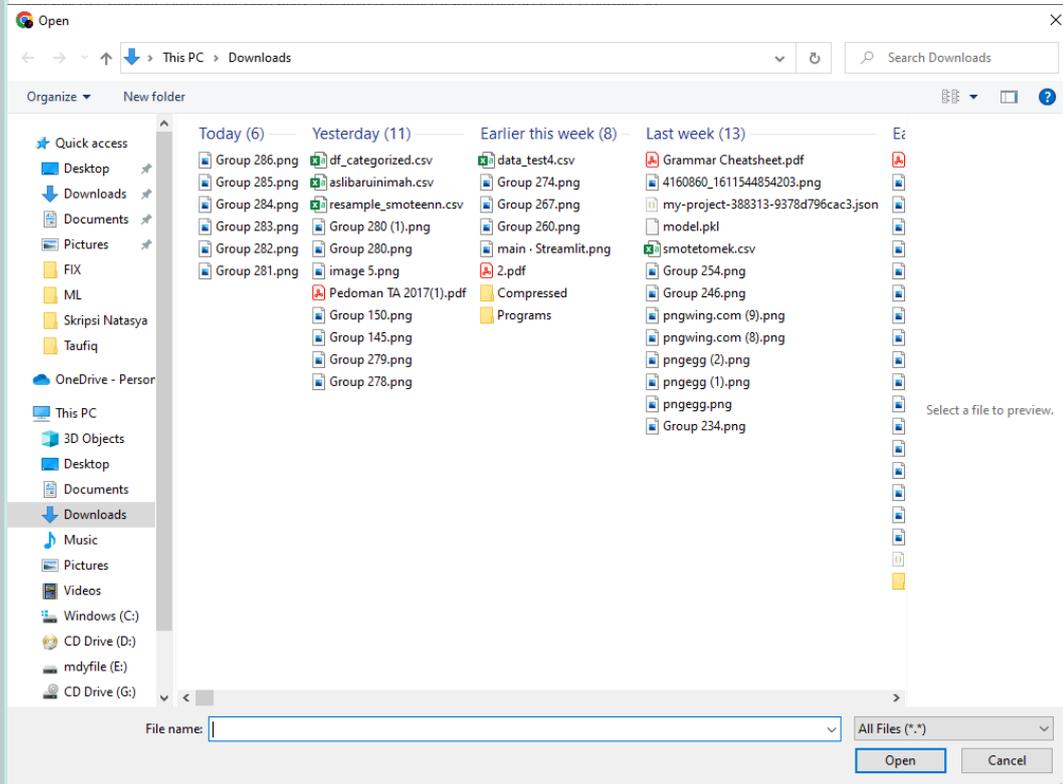
Detail file yang telah diupload

Gambar 3.71 Panduan Unggah File Pada Bucket – 1



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Sesaat setelah Bucket baru terbuat, akan muncul lama web seperti pada Gambar 3.71. *Upload file* dapat dilakukan dengan menekan tombol UPLOAD FILES. Setelah itu akan muncul *pop-up* untuk memilih *file*. Jika sudah menemukan *file* yang ingin di-*upload*, tekan tombol Open.



Gambar 3.72 Panduan Unggah File Pada Bucket – 2

7) Aktifkan Google Cloud Storage API

Untuk mengaktifkan API dan Service lain, perlu masuk ke laman terkait.

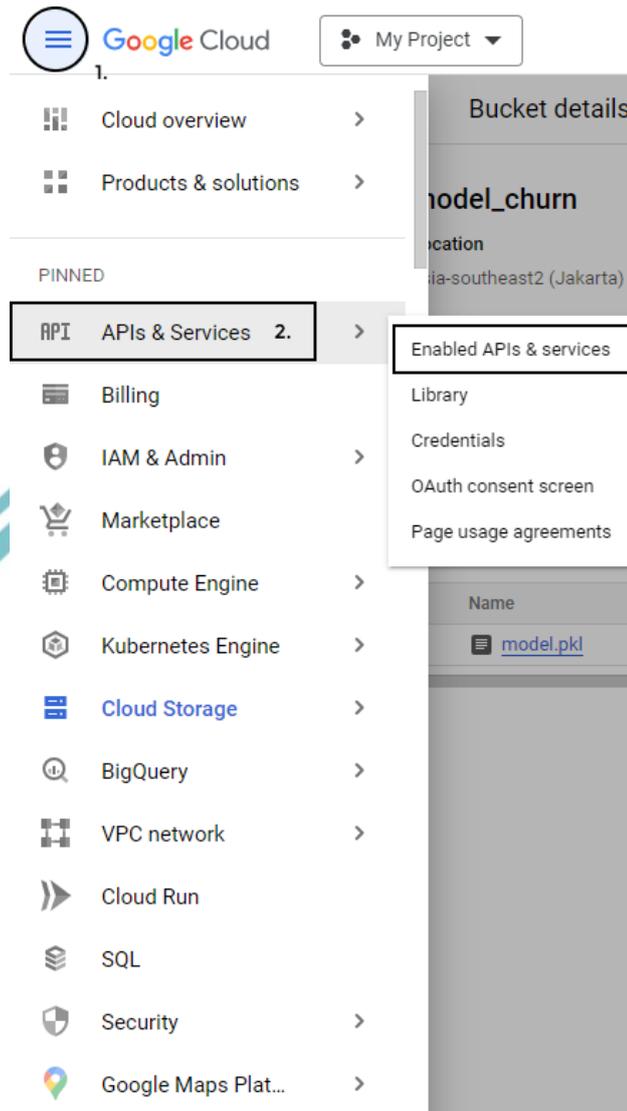
Gambar berikut menunjukkan cara untuk menuju laman tersebut.



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 3.73 Panduan Masuk Halaman Enabled APIs & services



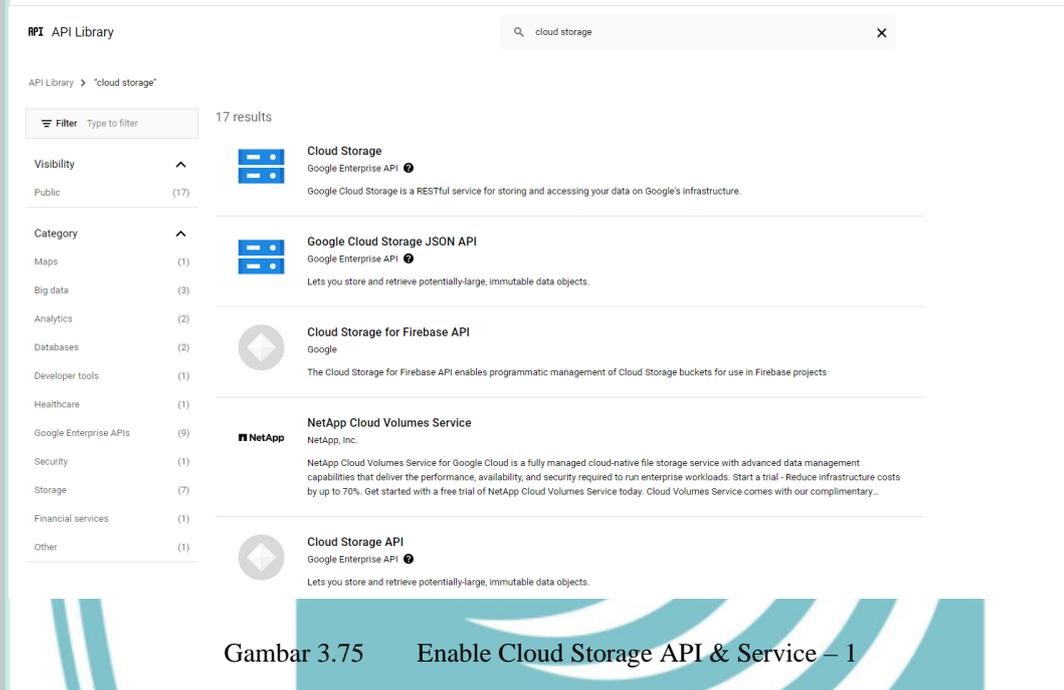
Gambar 3.74 Halaman APIs & Services



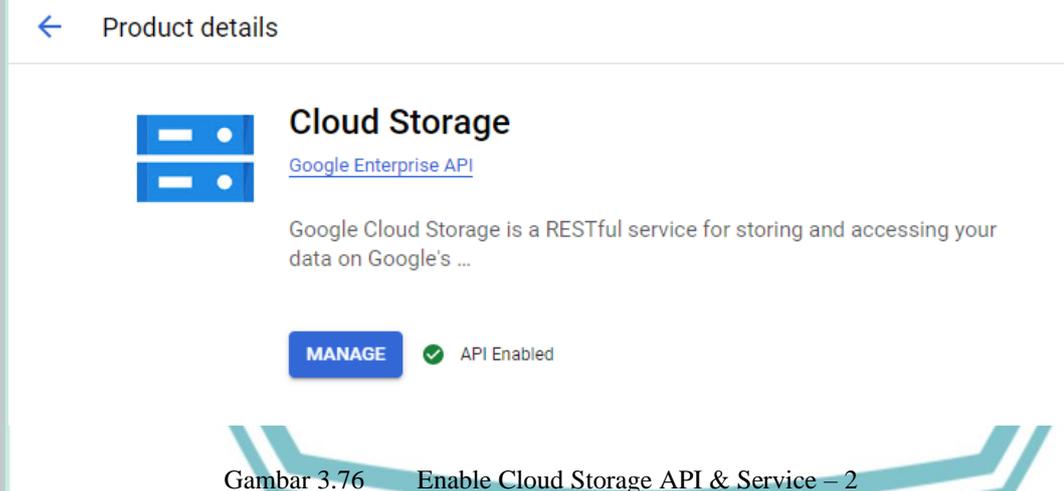
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Gambar 3.74 menunjukkan tampilan laman APIs&Services. Tekan tombol ENABLE APIS AND SERVICES untuk membuat API dan Service baru. Carilah Google Storage, lalu pilihlah Cloud Storage. Terakhir klik tombol MANAGE.



Gambar 3.75 Enable Cloud Storage API & Service – 1



Gambar 3.76 Enable Cloud Storage API & Service – 2

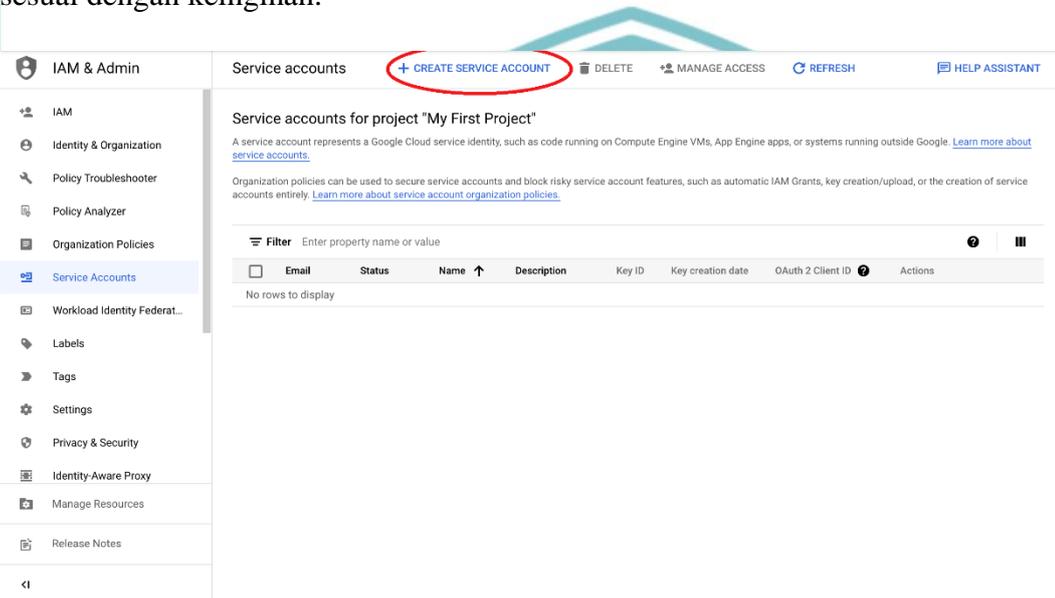
8) Membuat *service account* dan *file key*



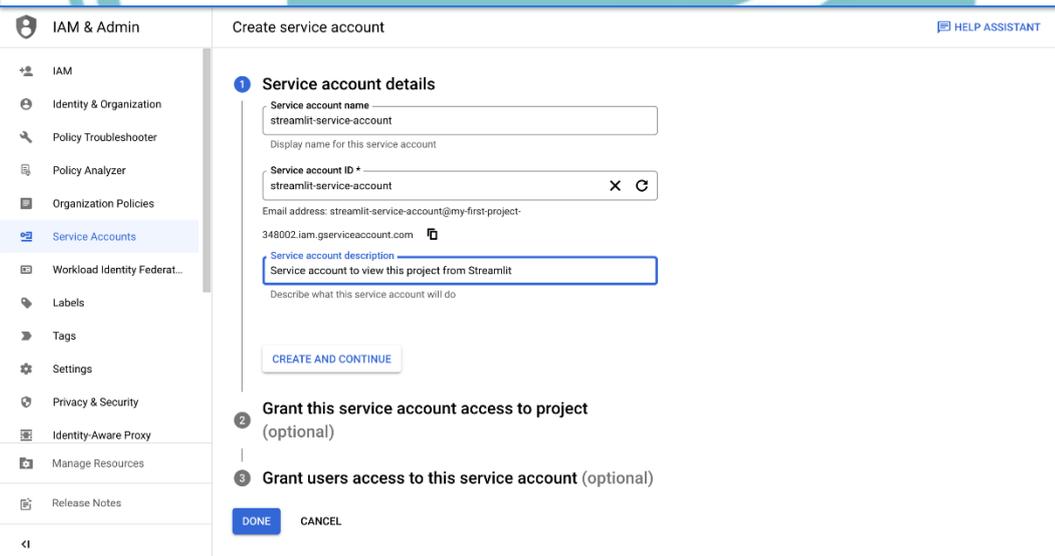
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Selanjutnya buatlah *service account* dan *file key*. Gambar 3.77 ini merupakan tampilan dari halaman Service accounts. Laman ini berada pada Menu IAM & Admin. Lalu pilih pada *sidebar* “Service Account”. Klik tombol CREATE SERVICE ACCOUNT untuk membuat *service account* baru. Lalu isikan detail sesuai dengan keinginan.



Gambar 3.77 Membuat Service Account – 1

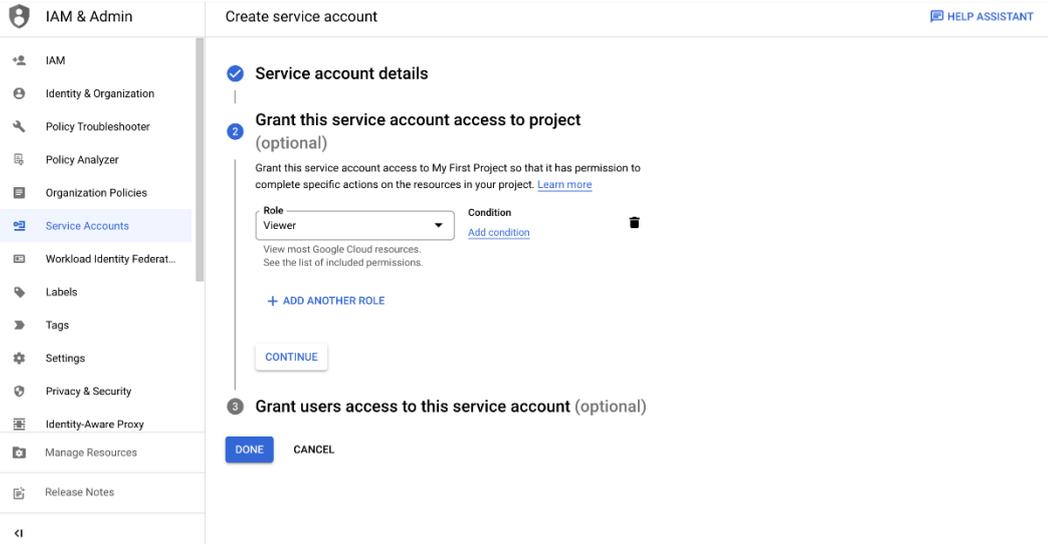


Gambar 3.78 Membuat Service Account – 2



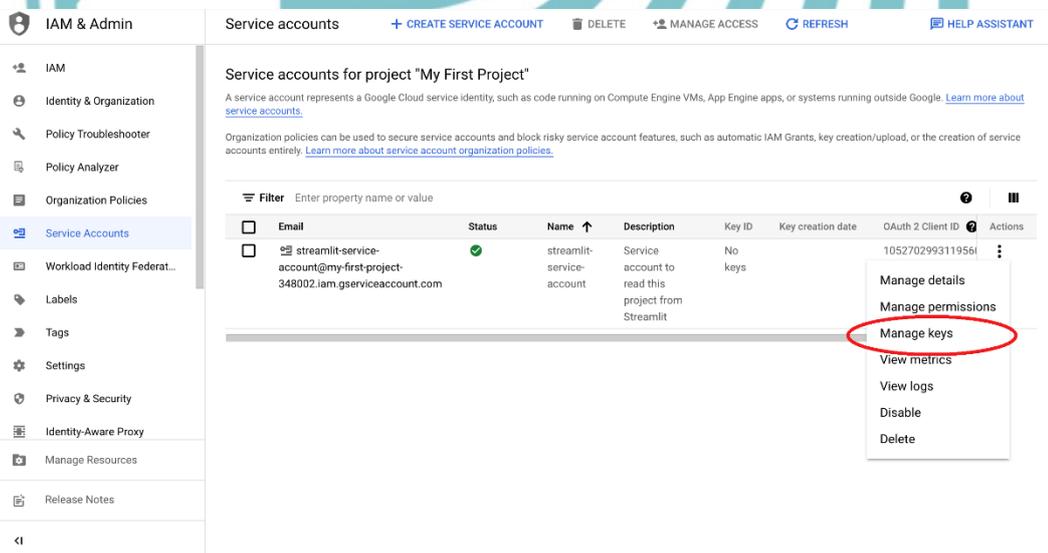
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 3.79 Membuat Service Account – 3

Atur *grant* untuk *service account* yang ingin dibuat. Lalu setelah itu, klik DONE.



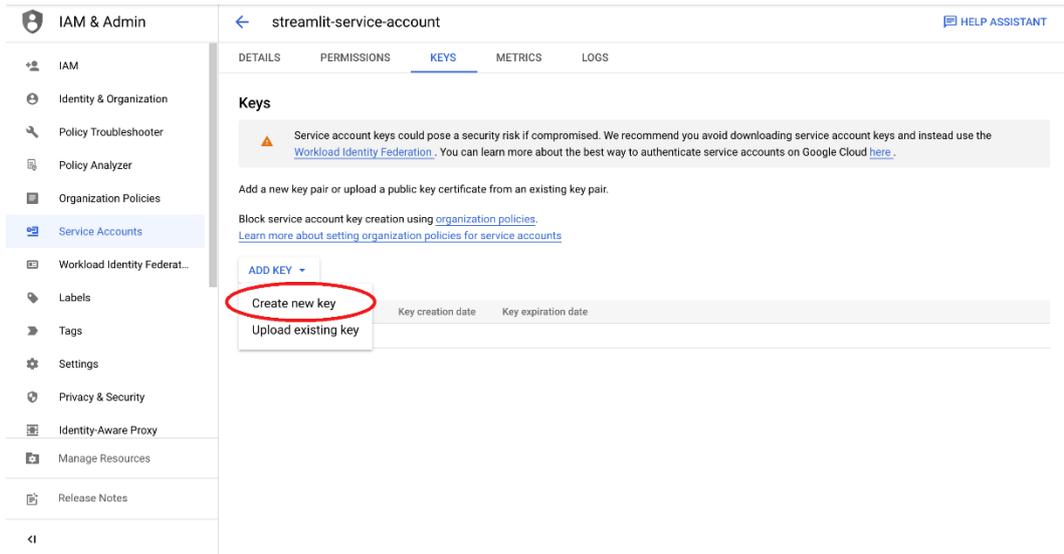
Gambar 3.80 Manage Key – 1

Kembali ke laman Service Account. Pilihlah *account service* yang telah dibuat. Pilih Manage Key untuk *managing* Key.



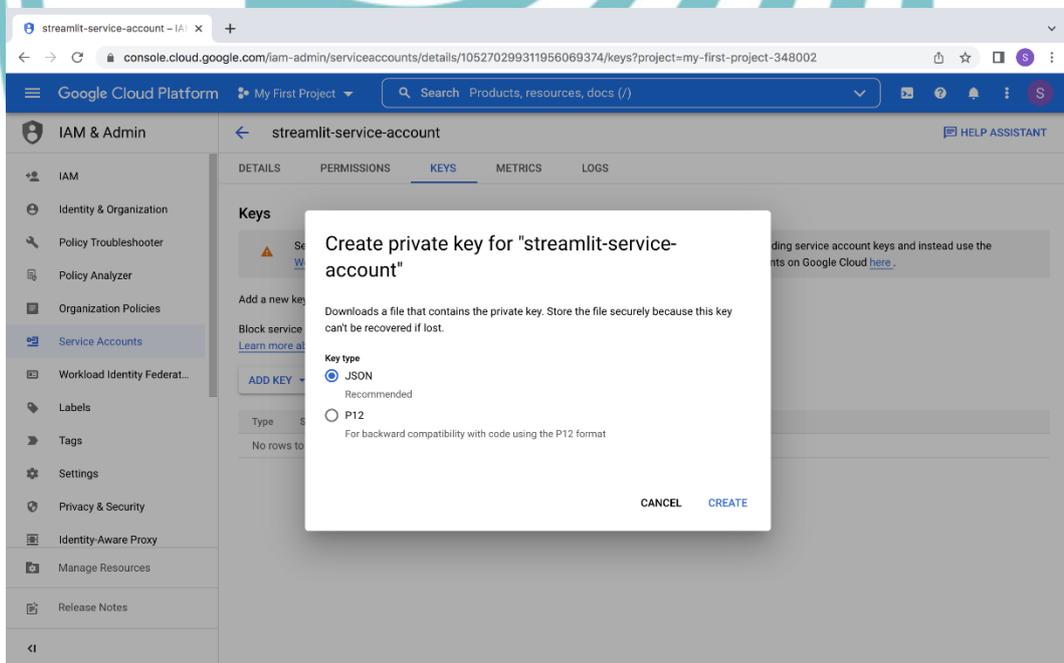
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 3.81 Manage Key – 2

Pilihlah “Create new key”.



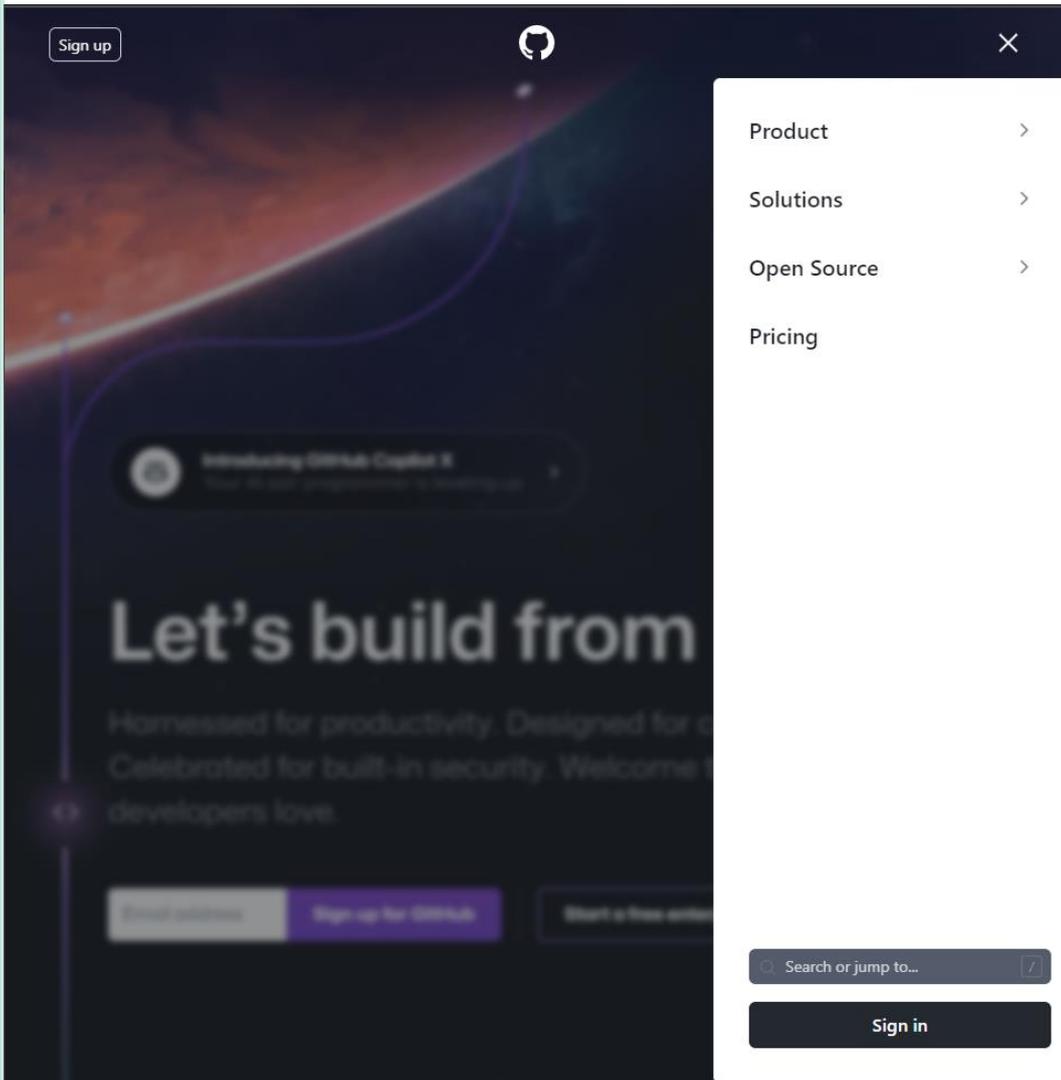
Gambar 3.82 Membuat Private Key JSON

Pilih tipe JSON. Setelah klik CREATE, *file* JSON akan otomatis ter-*download*. File JSON ini dibutuhkan untuk pengaturan pada *file* yang hendak di-*deploy* nantinya.



© Hak Cipta milik Politeknik Negeri Jakarta

- B. *Deploy Aplikasi*
 1) Upload *file* ke GitHub



Gambar 3.83 Proses Unggah File Di GitHub – 1

Buka *website* github.com. Lalu lakukan login dengan menekan tombol “Sign in”.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Sign in to GitHub

Username or email address

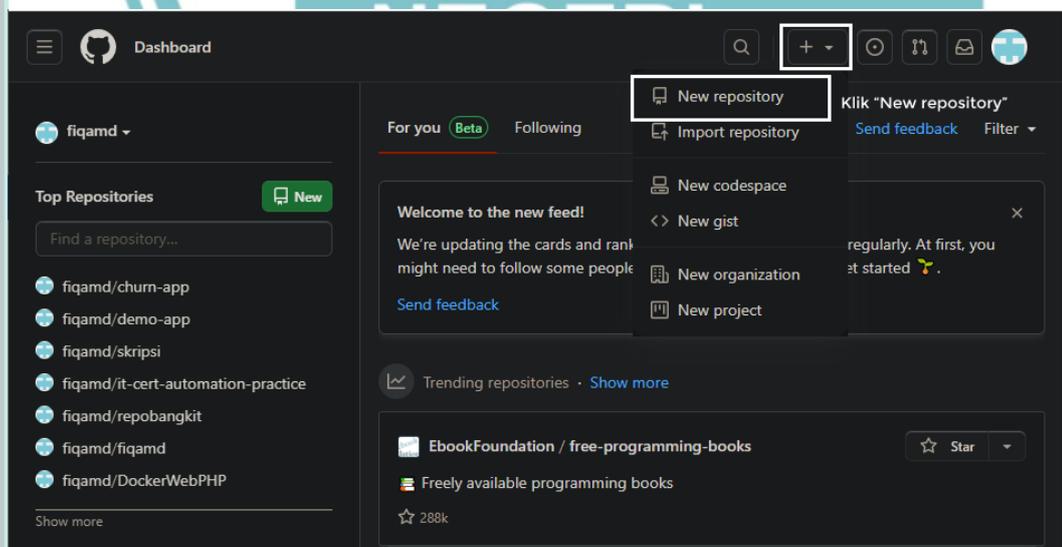
Password [Forgot password?](#)

Sign in

New to GitHub? [Create an account.](#)

Gambar 3.84 Proses Unggah File Di GitHub – 2

Isi form login sesuai dengan akun yang telah dibuat.



Gambar 3.85 Proses Unggah File Di GitHub – 1



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Buatlah Repository baru dengan menekan pilihan “New repository” yang ditunjukkan pada Gambar 3.85. Setelah pilihan “New repository” di tekan, maka akan muncul laman web yang terlihat seperti Gambar 3.86.

Gambar 3.86 Proses Unggah File Di GitHub – 1

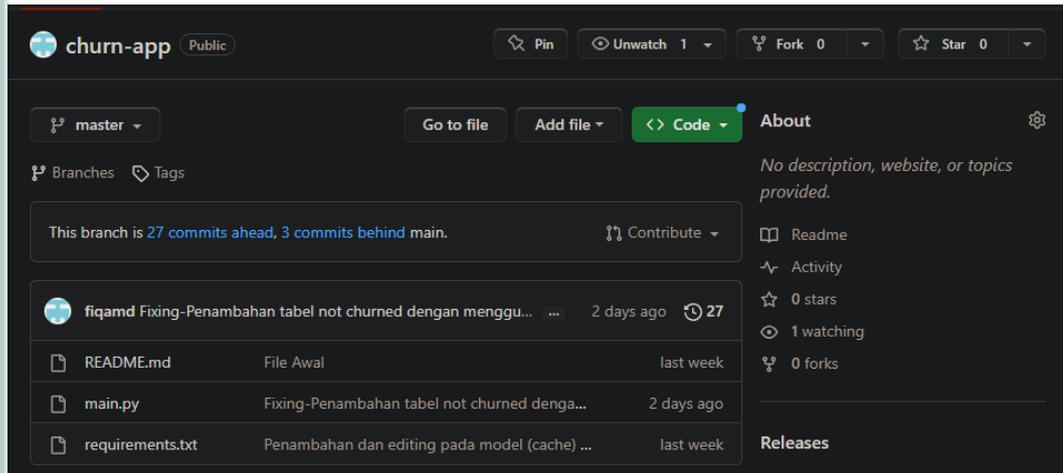
Isikan nama Repository, lalu buatlah Repository ini menjadi Public. Pengaturan tipe Repository ini harus dilakukan dengan *setting* Public, agar nanti pada saat *deploy* dengan menggunakan Streamlit tidak mengalami masalah.

Setelah Repository terbuat, *upload* seluruh *file* yang telah dibuat dengan klik tombol “Add file”.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

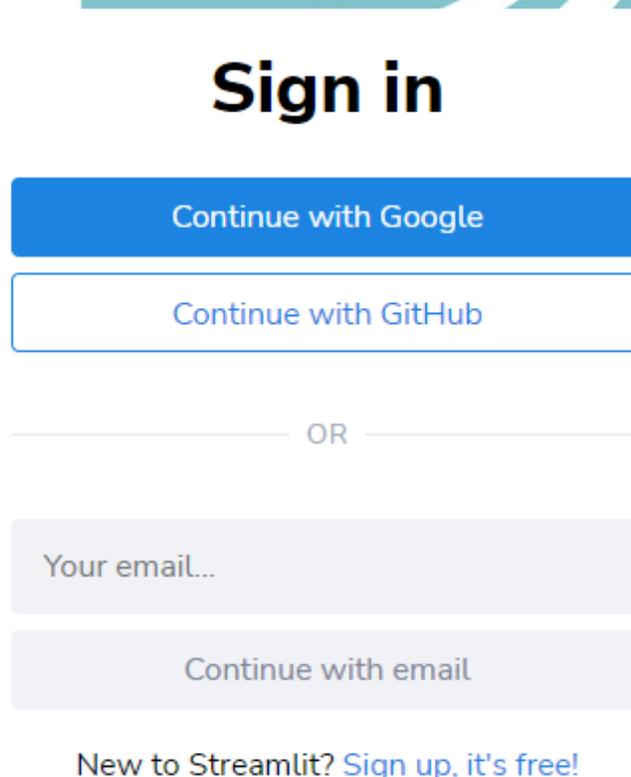


Gambar 3.87 Proses Unggah File Di GitHub – 1

Gambar 3.87 adalah tampilan ketika seluruh *file* telah di-*upload*.

2) Deploy dengan Streamlit

Akseslah *website* share.streamlit.io. Halaman *website* akan terlihat seperti Gambar 3.88.



Gambar 3.88 Proses Deploy Via Streamlit – 1

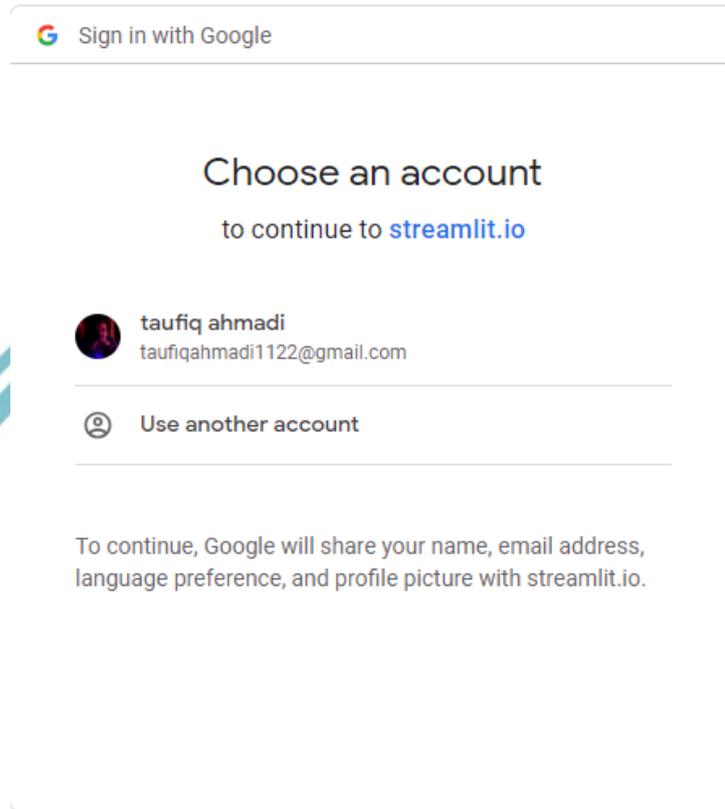


© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

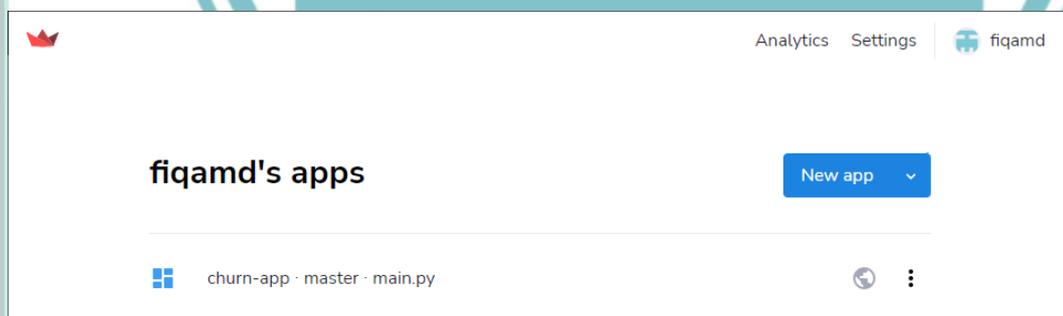
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Gambar 3.88 meminta untuk dilakukan login sebelum memulai *deploy* aplikasi. Di sini akan menggunakan akun GitHub agar mempermudah proses *deploy website* nantinya.



Gambar 3.89 Proses Deploy Via Streamlit – 2

Gambar 3.89 menunjukkan bahwa Streamlit harus melakukan login juga dengan menggunakan akun Google. Hal ini dilakukan agar Streamlit terkoneksi dengan Google Cloud Platform dan akun GitHub.



Gambar 3.90 Proses Deploy Via Streamlit – 3

Tampilan laman *website* dari *share.streamlit.io* terlihat seperti pada Gambar 3.90. Klik tombol “New app” untuk *deploy* aplikasi baru.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

[← Back](#)

Deploy an app

Repository [Paste GitHub URL](#)

fiqamd/repo

Branch

master

Main file path

streamlit_app.py

App URL (Optional)

.streamlit.app

[Advanced settings...](#)

[Deploy!](#)

Gambar 3.91 Proses Deploy Via Streamlit – 4

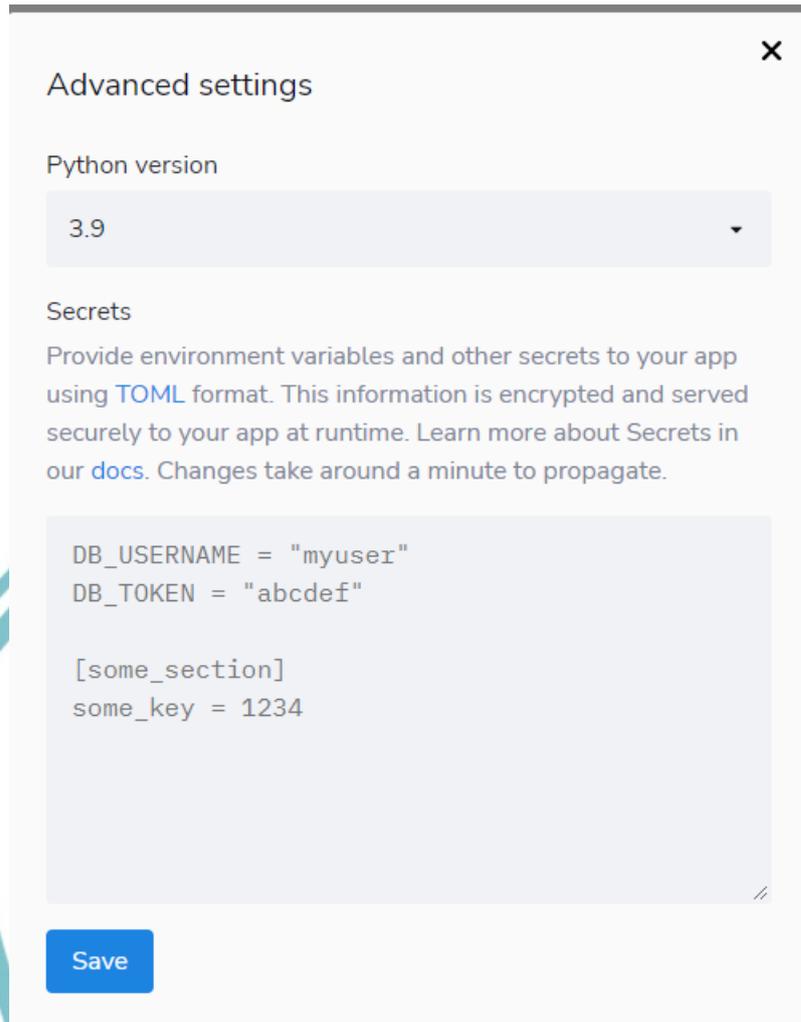
Setelah tombol “New app” diklik, maka akan muncul halaman *website* seperti pada Gambar 3.91. Isilah sesuai dengan Repository yang telah dibuat sebelumnya. Lalu klik “Advanced settings...”, maka akan tampil seperti pada Gambar 3.92 ini.



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 3.92 Proses Deploy Via Streamlit – 5

Isikan “Secrets” sesuai dengan format berikut.

```
# .streamlit/secrets.toml

[connections.gcs]
type = "service_account"
project_id = "xxx"
private_key_id = "xxx"
private_key = "xxx"
client_email = "xxx"
client_id = "xxx"
auth_uri = "https://accounts.google.com/o/oauth2/auth"
token_uri = "https://oauth2.googleapis.com/token"
auth_provider_x509_cert_url = "https://www.googleapis.com/oauth2/v1/certs"
client_x509_cert_url = "xxx"
```

Gambar 3.93 Proses Deploy Via Streamlit – 5



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Variabel yang bernilai “xxx” dapat diisi sesuai dengan *file key* JSON yang telah diunduh sebelumnya. Lakukanlah pengecekan pada *file* JSON, lalu isikan “Secrets” sesuai dengan *key* pada *file* JSON. Setelah dilakukan *input* data sesuai dengan *file* JSON pada “Secrets”, maka lakukanlah penyimpanan pengaturan “Secrets” dengan menekan tombol “Save”.

C. Penyesuaian Kode

1) Melakukan *Import library* tambahan

Penyesuaian kode dilakukan agar *website* dapat ter-*deploy* dan berjalan dengan baik. Dengan ini diperlukan beberapa *library* tambahan. *Library* tersebut akan digunakan untuk mengkoneksikan *file* aplikasi yang dibuat dengan Google Cloud Storage.

```
from google.cloud import storage
from st files connection import FilesConnection
```

Gambar 3.94 Import Library Tambahan

2) Melakukan penambahan pengaturan pada pemanggilan model *machine learning*

Gambar berikut merupakan pengaturan tambahan yang dibutuhkan agar *file* model *machine learning* yang sudah berada pada Google Cloud Storage dapat terkoneksi dan dapat berjalan dengan baik pada aplikasi yang dibuat.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
@st.cache(allow_output_mutation=True)
def load_model(bucket_name, model_path):
    storage_client = storage.Client()
    bucket = storage_client.bucket(bucket_name)
    blob = bucket.blob(model_path)

    # Download the model file to local file
    local_model_path = 'local_model.pkl'
    blob.download_to_file(open(local_model_path, 'wb'))

    # Load the model from local file
    model = joblib.load(local_model_path)

    return model

# # Anda dapat menggunakan fungsi ini untuk meload model Anda
model = load_model("model_churn", "model.pkl")
```

Gambar 3.95 Penyesuaian Pemanggilan Model *Machine Learning*

3.3 Mekanisme Pengujian Tugas Akhir

Mekanisme pengujian tugas akhir terdiri dari dua bagian utama, yaitu mekanisme pengujian model *Random Forest Classifier* dan mekanisme pengujian program aplikasi *website*.

3.3.1 Mekanisme Pengujian Model *Random Forest Classifier*

Pengujian Model dilakukan dengan 3 cara, yaitu menguji nilai-nilai metrik evaluasi, uji dengan data *dummy*, dan pengujian model dengan menggunakan data real.

1) Waktu Pengujian

Waktu pengujian dilaksanakan pada bulan Juli tahun 2023

2) Sumber / Subjek Penelitian

Mekanisme pengujian model *Random Forest Classifier* akan dilakukan dengan tiga pengujian. Pertama menguji metrik evaluasi, yaitu mengukur seberapa baik model *machine learning* yang telah dibuat. Kedua menguji dengan data *dummy*, yaitu untuk melihat apakah proses prediksi dapat berjalan dengan baik. Ketiga menguji dengan data *real*, yaitu untuk melihat sejauh mana tingkat ketepatan prediksi yang akan menggunakan data *real*.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

- 3) Metode Pengumpulan Data
Metode pengumpulan data untuk pengujian model *machine learning* dengan melakukan pengujian menggunakan beberapa sampel *dataset* yang telah dibuat.

3.3.1.1 Uji Nilai Metrik Evaluasi

Di tahap pengujian ini dilakukan dengan melihat hasil dari metrik-metrik evaluasi model, antara lain *Accuracy*, *F1 Score*, *Recall Score*, *ROC AUC Score*, dan *Precision Score*.



Gambar 3.96 Visualisasi Uji Metrik Evaluasi

Uji model ini akan menggunakan *dataset* bersih yang sudah disiapkan. Selanjutnya *dataset* tersebut akan melalui proses *training* dan *testing* dengan algoritma *Random Forest Classifier*. Setelah proses *training* dan *testing* selesai, maka akan keluar nilai-nilai metrik evaluasi.

3.3.1.2 Uji Dengan Data *Dummy*

Di pengujian ini akan melihat bagaimana *output* atau hasil prediksi yang diberikan oleh model *machine learning* dengan algoritma *Random Forest Classifier* yang sudah dibuat. Pembuatan data *dummy*, dilakukan dengan *blind sampling*. Hal ini dilakukan karena data yang dibuat ini adalah data *dummy* dan belum mengetahui hasil akhirnya akan seperti apa. Nantinya akan membuat 3 *file* random data *.csv*



yang akan diprediksi oleh model dengan masing-masing data memiliki skenario berbeda. Detail dari ketiga *file* dapat disimak sebagai berikut.

A. *Dataset 1*

Dataset 1 ini dibuat dengan komposisi sebagai berikut:

- *Dataset* akan memuat data churn sebanyak 1.475 dan data random sebanyak 47.692 atau dengan rasio sekitar 3% : 97%.
- Data random dibuat dengan mengambil 15 jenis *unique value* dari kolom *Area Name, Plan, Tv Plan, dan Advance Promo*. Selain itu, data juga dibuat dengan nilai kolom *Complaint by Customer Service, Complaint by Email, Complaint by Social Media, Complaint by Telegram, Complaint by Whatsapp, dan Complaint by WIC* masing-masing dengan angka acak dari 1-200.

B. *Dataset 2*

Dataset 2 ini dibuat dengan komposisi sebagai berikut:

- *Dataset* akan memuat data churn sebanyak 1.475 dan data random sebanyak 47.692 atau dengan rasio sekitar 3% : 97%.
- Data random dibuat dengan mengambil semua jenis *unique value* dari kolom *Area Name, Plan, Tv Plan, dan Advance Promo*. Selain itu, data juga dibuat dengan nilai kolom *Complaint by Customer Service, Complaint by Email, Complaint by Social Media, Complaint by Telegram, Complaint by Whatsapp, dan Complaint by WIC* masing-masing dengan angka acak dari 1-200.

C. *Dataset 3*

Dataset 3 ini dibuat dengan mengambil semua jenis *unique value* dari kolom *Area Name, Plan, Tv Plan, dan Advance Promo*. Selain itu, data juga dibuat dengan nilai kolom *Complaint by Customer Service, Complaint by Email, Complaint by Social Media, Complaint by Telegram, Complaint by Whatsapp, dan Complaint by WIC* masing-masing dengan angka acak dari 1-200. *Dataset 3* ini akan memuat data sebanyak 140.000.

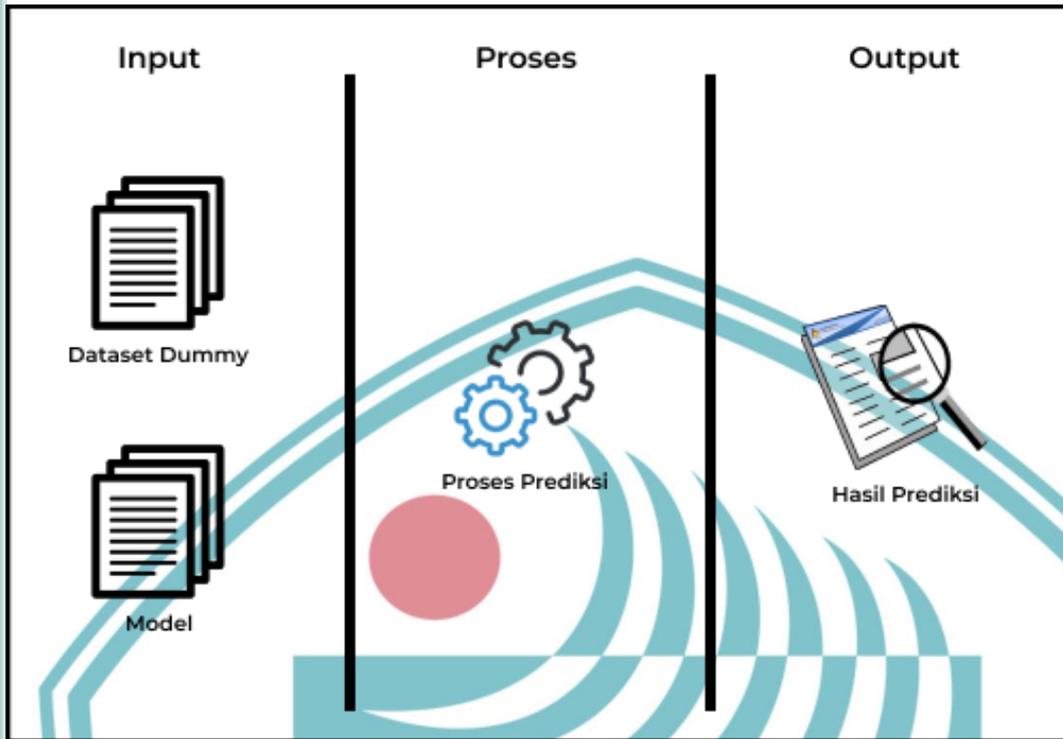
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, pennisan karya ilmiah, pennisan laporan, pennisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

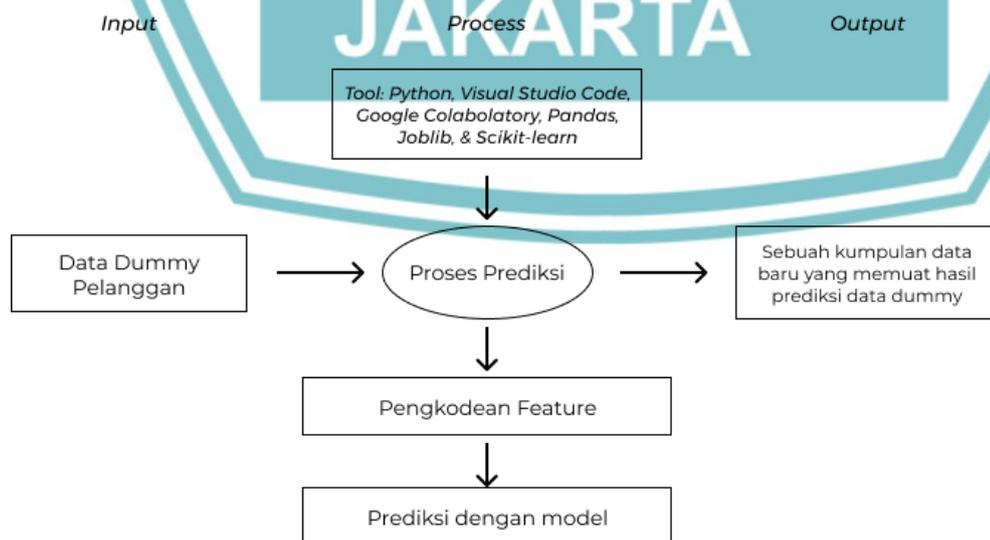
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 3.97 Visualisasi Uji Dengan Data Dummy

Uji dengan data *dummy* ini memerlukan model dan *dataset* yang akan diprediksi. Nantinya *dataset dummy* ini akan diprediksi dengan menggunakan model yang sudah disiapkan. Langkah terakhir, melihat bagaimana *output* dari masing-masing *file* seperti apa.

Adapun blok diagram untuk mekanisme uji model dengan data *dummy* sebagai berikut.

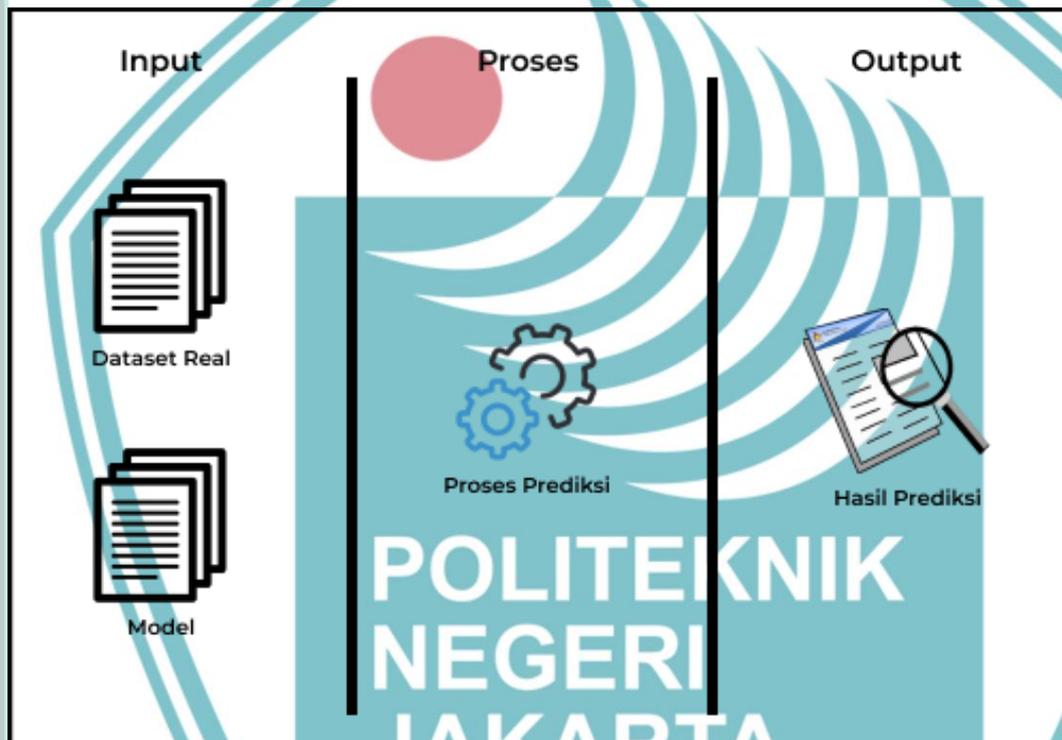


Gambar 3.98 Blok Diagram Mekanisme Uji Data Dummy



3.3.1.3 Uji Data *Real*

Di pengujian ini akan dilihat *output* yang dihasilkan oleh sistem prediksi churn yang sudah dibuat. Pada pengujian kali ini akan digunakan *dataset* yang diambil dari *dataset* asli dari perusahaan. Tujuan uji data *real* ini sendiri dilakukan untuk melihat seberapa sesuai *output* yang diberikan oleh sistem prediksi churn yang telah dibuat dengan hasil akhir yang ada pada *dataset* perusahaan. Data *real* yang digunakan akan memuat data churn sebanyak 1.475 yang diambil dari data pelanggan yang sudah churn pada perusahaan.



Gambar 3.99 Visualisasi Uji Data *Real*

Uji dengan data *real* ini memerlukan model dan *dataset* yang akan diprediksi. Nantinya *dataset* yang diambil dari *dataset* perusahaan ini akan diprediksi dengan menggunakan model yang sudah disiapkan. Langkah terakhir yang dilakukan, yaitu membandingkan hasil data yang telah diprediksi dengan data asli dari perusahaan.

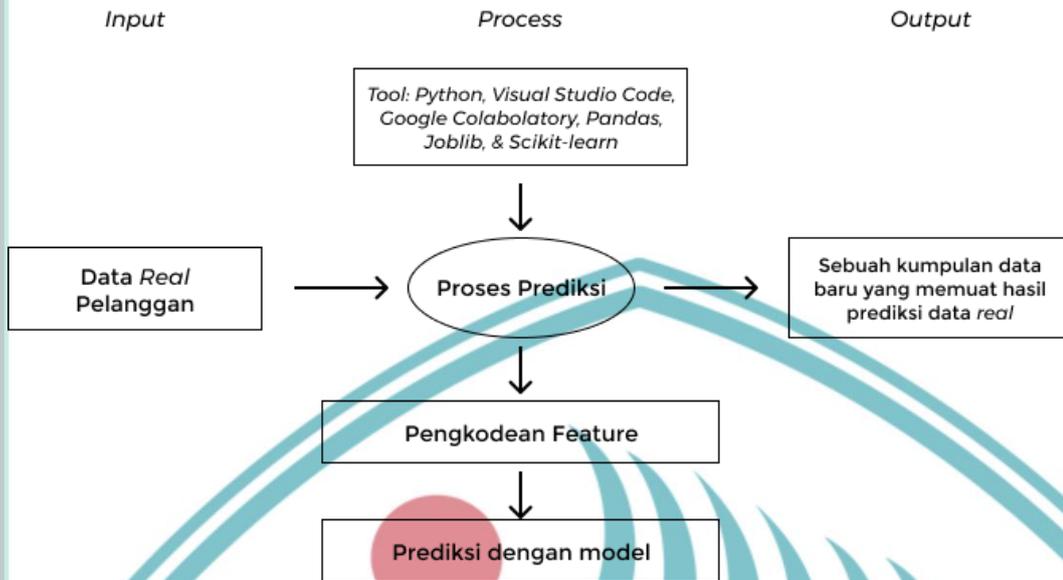
Adapun blok diagram untuk mekanisme uji model dengan data *real* sebagai berikut.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 3.100 Diagram Blok Uji Data Real

3.3.2 Mekanisme Pengujian *Generate Churn Report*

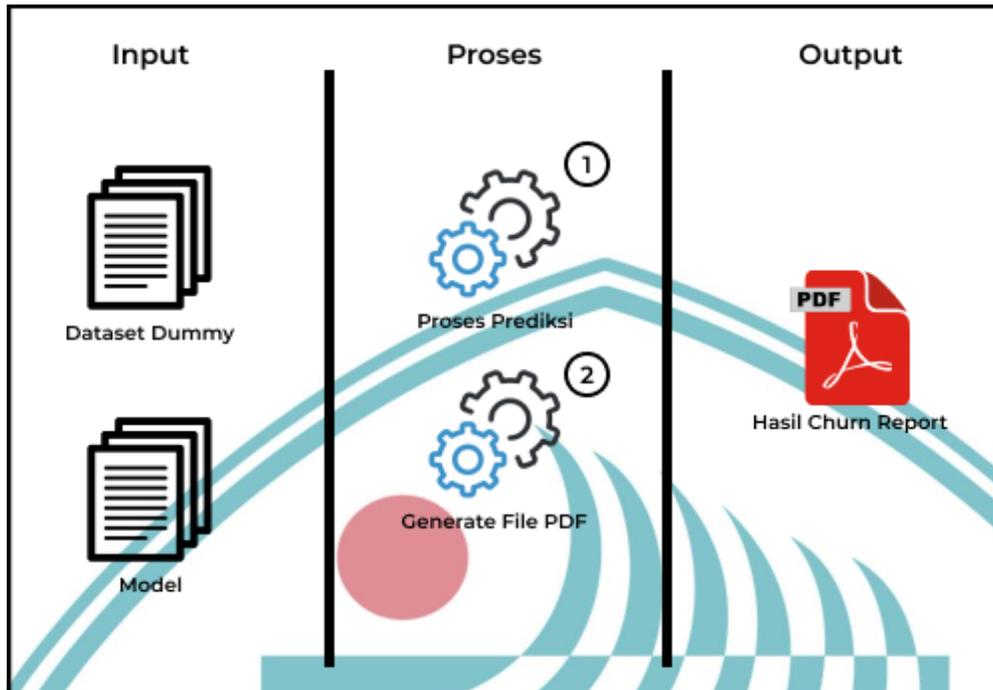
Pengujian *generate churn report* dilakukan dengan cara menjalankan suatu tombol yang nantinya akan memunculkan tombol lainnya dan tombol tersebut akan otomatis melakukan pengunduhan terhadap *file* PDF yang sudah ter-*generate* sebelumnya.

- 1) Waktu Pengujian
Waktu pengujian dilaksanakan pada bulan Juli tahun 2023
- 2) Sumber / Subjek Penelitian
Mekanisme pengujian *generate churn report* dilakukan pada *website* prediksi churn yang telah dibuat. Pengujian ini dilakukan untuk melihat apakah proses *generate file* PDF untuk *reporting* hasil prediksi churn dapat dibuat dan diunduh dengan baik pada *website* yang telah dibuat.
- 3) Metode Pengumpulan Data
Metode pengumpulan data untuk pengujian *generate churn report* dengan melakukan pengujian menggunakan tiga *dataset dummy* yang telah dibuat.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 3.101 Visualisasi Pengujian Generate Churn Report

Uji *generate churn report* ini membutuhkan *input* berupa *file model* dan *dataset dummy*. *File model* ini merupakan *model customer churn prediction* yang telah dibuat, sementara itu *dataset dummy* merupakan *dataset* yang akan digunakan selama proses pengujian ini berlangsung. Nantinya data *dummy* ini akan diproses dengan *model customer churn prediction* yang selanjutnya akan *generating file PDF* yang akan memuat hasil dari seluruh grafik yang telah divisualisasikan juga pada *website*. *File churn report* ini nantinya akan dapat diunduh dengan menekan sebuah tombol.

3.3.3 Mekanisme Pengujian Website Prediksi Churn Pelanggan

Tahap pengujian *website* prediksi churn pelanggan bertujuan untuk memastikan apakah *website* yang telah dibuat sudah berjalan sesuai dengan spesifikasi dan kebutuhan. Berikut ini merupakan mekanisme pengujian *website* prediksi churn pelanggan.

- 1) Waktu Pengujian
Waktu pengujian dilaksanakan pada bulan Juli tahun 2023
- 2) Sumber / Subjek Penelitian
Mekanisme pengujian *website* pada *website* prediksi churn pelanggan dilakukan berdasarkan *User Acceptance Test (UAT)*



© Hak Cipta milik Politeknik Negeri Jakarta

3) Metode Pengumpulan Data

Metode pengumpulan data yang digunakan pada penelitian ini adalah metode kuesioner. Metode kuesioner dilakukan dengan menyebarkan sejumlah pertanyaan kepada satu atau lebih pengguna.

Contract Acceptance Testing (CAT) merupakan proses pengujian yang dilakukan oleh pengguna dengan maksud untuk menghasilkan bukti bahwa sistem *website* yang dikembangkan dapat diterima atau ditolak oleh pengguna. Pengujian CAT dilakukan dengan mengajukan beberapa pertanyaan terhadap para pengguna, dimana pada pengujian ini melibatkan 5 *user* admin dan 11 *user* officer. Hasil *Contract Acceptance Test* (CAT) dinilai berdasarkan 5 kategori, yaitu Sangat Setuju (SS), Sesuai (S), Kurang Sesuai (KS), Tidak Sesuai (TS) dan Tidak Menjawab (TM).

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



BAB IV PEMBAHASAN

Bab ini merupakan tahap terakhir dalam proses pembuatan skripsi. Pembahasan pada bab ini meliputi proses pengujian model *Random Forest Classifier* dan pengujian *website* prediksi churn pelanggan. Kedua proses pengujian tersebut memiliki tujuan untuk memastikan bahwa sistem dapat berjalan dengan baik dan menghasilkan saran yang dapat diterima oleh penggunanya.

Pengujian model *machine learning* dilakukan dengan 3 cara, yaitu pengujian terhadap nilai metrik evaluasi yang dimiliki oleh model tersebut, pengujian model dengan menggunakan data *dummy*, dan pengujian model dengan data *real*. Pengujian terhadap nilai metrik evaluasi yang dimiliki, bertujuan untuk melihat kualitas dari model yang telah dibuat. Pengujian ini akan melihat nilai metrik evaluasi *Accuracy*, *F1 Score*, *Recall Score*, *ROC AUC Score*, dan *Precision Score*. Pengujian model dengan menggunakan data *dummy* bertujuan untuk melihat apakah model yang sudah dibuat dapat memprediksi dan berjalan dengan baik dengan data *input* berupa 3 data *random*. Pengujian model dengan menggunakan data *real* bertujuan untuk melihat seberapa akurat dari model yang telah dibuat dengan pengimplementasian langsung menggunakan data *real* perusahaan yang memuat data churn.

Sedangkan pada pengujian *website* prediksi churn pelanggan dilakukan dengan metode *User Acceptance Test (UAT)* untuk memastikan bahwa *website* yang dibuat dapat diterima dan sesuai dengan kebutuhan dari pengguna.

Pengujian dilakukan berdasarkan informasi berikut:

1. Waktu : Juli 2023
2. Pelaksana : Taufiq Ahmadi
3. Instruktur : Asri Wulandari, S.T., M.T.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

4.1 Pengujian Model *Machine Learning*

Pengujian model *Machine Learning* dilakukan dengan 3 cara, yaitu pengujian terhadap nilai metrik evaluasi yang dimiliki oleh model tersebut, pengujian model dengan menggunakan data *dummy*, dan pengujian model dengan data *real*.

4.1.1 Pengujian Nilai Metrik Evaluasi

4.1.1.1 Deskripsi Pengujian

Pengujian nilai metrik evaluasi dari model *Random Forest Classifier* yang telah dibuat bertujuan untuk melihat performansi dari model yang telah terbuat yang tercatat di atas kertas. Hal ini perlu dilakukan sebelum dilakukan penetapan model karena untuk memberikan bayangan akan kualitas model yang telah dibuat.

4.1.1.2 Prosedur Pengujian

Berikut adalah prosedur pengujian nilai metrik evaluasi model dengan melihat *Accuracy Score*, *F1 Score*, *Recall*, *Precision Score*, dan *ROC AUC Score*:

1. Menyiapkan *dataset* yang sudah siap pakai
2. Mengacak *dataset* latih dan uji
3. Melatih dan menguji model dengan menggunakan algoritma *Random Forest Classifier*
4. Menganalisis metrik-metrik evaluasi yang dihasilkan oleh model yang telah dibuat
5. Melakukan pengulangan langkah (2) sampai (4) hingga 5 kali percobaan
6. Mengambil nilai rata-rata yang dihasilkan dari 5 kali percobaan yang telah dilakukan.

Pengujian ini dilakukan dengan cara *training* dan *testing dataset* yang sudah disiapkan dengan algoritma *machine learning* bernama *Random Forest Classifier*. *Output* dari *training* dan *testing* ini selanjutnya akan dianalisis. Data ini akan berupa hasil dari metrik-metrik evaluasi model, antara lain *Accuracy*, *F1 Score*, *Recall Score*, *ROC AUC Score*, dan *Precision Score*. Selanjutnya akan metrik-metrik evaluasi tersebut akan dianalisis berdasarkan skor yang diperoleh masing-masing metrik evaluasi. Berdasarkan hasil analisis, akan disimpulkan apakah model ini sudah baik dan dapat diterapkan pada prediksi churn pelanggan yang akan dibuat.



4.1.1.3 Data Hasil Pengujian

Pengujian ini dilakukan dengan cara *training* dan *testing dataset* yang sudah disiapkan dengan algoritma *machine learning* bernama *Random Forest Classifier*. *Output* dari *training* dan *testing* berupa metrik evaluasi. Adapun metrik evaluasi yang dimaksud dari hasil *training* dan *testing* algoritma sebagai berikut.

1) *Accuracy Score*

Setelah dilakukan *training* dan *testing* dengan menggunakan algoritma *machine learning Random Forest Classifier* akan muncul salah satu metrik evaluasi bernama *Accuracy*. Sesuai dengan namanya, metrik evaluasi ini akan mengukur seberapa akurat model *machine learning* yang telah dibuat. Secara sistematis, *Accuracy score* dihitung dengan membandingkan jumlah prediksi yang benar (*true predictions*) dengan jumlah total sampel yang dinilai. Adapun formula untuk menghitung *Accuracy Score* sebagaimana yang disebutkan pada formula (2.1).

2) *F1 Score*

Setelah dilakukan *training* dan *testing* dengan menggunakan algoritma *machine learning Random Forest Classifier* akan muncul salah satu metrik evaluasi bernama *F1 Score*. *F1 Score* ini memberikan informasi tentang keseimbangan antara *precision* dan *Recall* dari model. *F1 Score* dihitung dengan menggabungkan *precision* dan *Recall* menggunakan *harmonic mean*. *Precision* mengukur seberapa akurat model dalam mengidentifikasi pelanggan yang benar-benar churn, sementara *Recall* mengukur seberapa baik model dalam menemukan semua kasus pelanggan yang sebenarnya churn. Oleh karena itu, *F1 Score* dapat memberikan gambaran yang lebih baik tentang kinerja model di bawah ketidakseimbangan data. Formula untuk menghitung *F1 Score* sebagaimana yang disebutkan pada formula (2.2).

3) *Recall Score*

Setelah dilakukan *training* dan *testing* dengan menggunakan algoritma *machine learning Random Forest Classifier* akan muncul salah satu metrik evaluasi bernama *Recall Score*. *Recall Score* ini mengukur seberapa baik model dapat menemukan semua kasus positif (sebagai contoh, pelanggan yang benar-benar churn) dari keseluruhan jumlah kasus positif yang ada pada data. *Recall*

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



© Hak Cipta milik Politeknik Negeri Jakarta

memberikan informasi tentang kemampuan model untuk mengidentifikasi sebagian besar pelanggan yang sebenarnya churn. Singkatnya, *Recall* mengukur proporsi kasus positif yang berhasil diidentifikasi oleh model dari keseluruhan jumlah kasus positif yang ada pada data. Formula untuk menghitung *Recall Score* sebagaimana yang disebutkan pada formula (2.3).

4) *ROC AUC Score*

Setelah dilakukan *training* dan *testing* dengan menggunakan algoritma *machine learning Random Forest Classifier* akan muncul salah satu metrik evaluasi bernama *ROC AUC Score*. *ROC (Receiver Operating Characteristic)* dan *AUC (Area Under the Curve)* ialah metrik evaluasi yang digunakan untuk mengukur kinerja model *machine learning* pada berbagai ambang batas (*threshold*). *ROC AUC* digunakan untuk memahami seberapa baik model dapat membedakan antara dua kelas target (positif dan negatif) dalam suatu masalah klasifikasi. Formula untuk menghitung *Recall Score* sebagaimana yang disebutkan pada formula (2.4) dan (2.5).

5) *Precision Score*

Setelah dilakukan *training* dan *testing* dengan menggunakan algoritma *machine learning Random Forest Classifier* akan muncul salah satu metrik evaluasi bernama *Precision Score*. *Precision Score* ialah metrik evaluasi yang digunakan untuk mengukur mengukur sejauh mana model klasifikasi benar dalam mengidentifikasi dan memprediksi data sebagai kelas positif dari semua prediksi yang dibuat sebagai kelas positif. Nilai *Precision* berkisar dari 0 hingga 1, di mana 1 menunjukkan bahwa model membuat prediksi positif yang benar-benar sempurna dan 0 menunjukkan bahwa model sama sekali tidak memiliki prediksi positif yang benar. Semakin tinggi nilai *Precision*, semakin baik model dapat mengidentifikasi data dengan benar sebagai kelas positif. Formula untuk menghitung *Precision Score* sebagaimana yang disebutkan pada formula (2.4) dan (2.5).

Setelah dilakukan proses *training* dan *testing dataset* dengan algoritma *Random Forest Classifier* sebanyak 5 kali uji coba, dihasilkan data pengujian sebagai berikut:

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Uji 1		
	Metric	Value
0	Accuracy	0.946158
1	F1 Score	0.944044
2	Recall	0.950900
3	Precision	0.937285
4	ROC AUC	0.946361

Gambar 4.1 Hasil Uji Nilai Metrik Evaluasi 1

Uji 2		
	Metric	Value
0	Accuracy	0.946140
1	F1 Score	0.943664
2	Recall	0.950646
3	Precision	0.936784
4	ROC AUC	0.946359

Gambar 4.2 Hasil Uji Nilai Metrik Evaluasi 2

Uji 3		
	Metric	Value
0	Accuracy	0.946113
1	F1 Score	0.944013
2	Recall	0.953677
3	Precision	0.934543
4	ROC AUC	0.946454

Gambar 4.3 Hasil Uji Nilai Metrik Evaluasi 3

Uji 4		
	Metric	Value
0	Accuracy	0.946397
1	F1 Score	0.944765
2	Recall	0.954012
3	Precision	0.935695
4	ROC AUC	0.946683

Gambar 4.4 Hasil Uji Nilai Metrik Evaluasi 4

Uji 5		
	Metric	Value
0	Accuracy	0.946788
1	F1 Score	0.944814
2	Recall	0.950732
3	Precision	0.938968
4	ROC AUC	0.946946

Gambar 4.5 Hasil Uji Nilai Metrik Evaluasi 5



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Setelah mendapatkan nilai-nilai metrik evaluasi dari 5 kali pengujian, selanjutnya dilakukan perhitungan untuk mencari nilai rata-rata yang didapatkan oleh model yang telah dibuat. Data tersebut dapat dilihat pada tabel di bawah ini:

Tabel 4.1 Data Pengujian Nilai Metrik Evaluasi

Uji Coba	<i>Accuracy Score</i>	<i>F1 Score</i>	<i>Recall Score</i>	<i>Precision Score</i>	<i>ROC AUC Score</i>
Uji Coba 1	0,946158	0,944044	0,9509	0,937285	0,946361
Uji Coba 2	0,94614	0,943664	0,950646	0,936784	0,946359
Uji Coba 3	0,946592	0,944773	0,950803	0,93882	0,946751
Uji Coba 4	0,946397	0,944765	0,954012	0,935695	0,946683
Uji Coba 5	0,946788	0,944814	0,950732	0,938968	0,946946
Rata-Rata	0,946419	0,944324	0,95077	0,937964	0,946604

4.1.1.4 Analisa Hasil Pengujian

Berdasarkan hasil pengujian nilai metrik evaluasi pada Tabel 4.1, dapat dilihat bahwa nilai rata-rata dari *Accuracy Score* mendapatkan nilai sebesar 0,946419 atau jika diubah menjadi persentase sebesar 94,6419% yang dibulatkan menjadi 94,6%, *F1 Score* mendapatkan nilai sebesar 0,944324 atau jika diubah menjadi persentase sebesar 94,4324% yang dibulatkan menjadi 94,4%, *Recall Score* mendapatkan nilai sebesar 0,95077 atau jika diubah menjadi persentase sebesar 95,077% yang dibulatkan menjadi 95%, *Precision Score* mendapatkan nilai sebesar 0,937964 atau jika diubah menjadi persentase sebesar 93,7964% yang dibulatkan menjadi 93,8%, dan metrik evaluasi ROC AUC mendapatkan nilai sebesar 0,946604 atau jika diubah menjadi persentase sebesar 94,6604% yang dibulatkan menjadi 94,7%.

Dari penjabaran nilai rata-rata pada setiap metrik evaluasi dapat dikatakan bahwa model sudah dapat mengukur kemampuan model untuk mengklasifikasikan data dengan benar, dengan metrik evaluasi *Accuracy Score* yang menjadi acuannya. Model yang dibuat juga sudah dikatakan mampu mengukur dengan baik kemampuan model dalam mengklasifikasikan kelas positif dengan benar sekaligus menghindari kesalahan dalam mengklasifikasikan kelas negatif, dengan metrik *F1*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Score yang menjadi acuannya. Model juga dikatakan baik dalam hal mengukur jumlah positif aktual yang diidentifikasi model sebagai kelas positif, dengan metrik evaluasi *Recall Score* yang menjadi acuannya. Model dapat mengidentifikasi data dengan benar sebagai kelas positif, dengan acuan metrik evaluasi *Precision Score* di angka 93,8% dari nilai maksimal 100% atau 1. Terakhir, model memiliki kinerja yang baik pada beberapa ambang batas klasifikasi dan memprediksi kelas positif dan negatif dengan tingkat kebenarannya di angka 94,7%, dengan *ROC-AUC Score* yang menjadi acuannya.

4.1.2 Pengujian Dengan Data *Dummy*

4.1.2.1 Deskripsi Pengujian

Pengujian model yang telah dibuat dengan menggunakan data *dummy* bertujuan untuk melihat hasil dari proses prediksi churn pelanggan. Selain itu, hal ini dilakukan untuk meninjau sistem prediksi churn pelanggan yang telah dibuat dapat berjalan dengan baik atau tidak.

4.1.2.2 Prosedur Pengujian

Berikut adalah prosedur pengujian model dengan menggunakan data *dummy* sebagai *dataset* ujinya.

1. Menyiapkan 3 jenis *dataset dummy*. Ketentuan dalam membuat setiap *dataset* sebagai berikut:

- a. *Dataset 1*:

Dataset 1 ini dibuat dengan komposisi sebagai berikut:

- *Dataset* akan memuat data churn sebanyak 1.475 dan data random sebanyak 47.692 atau dengan rasio sekitar 3% : 97%.
- Data random dibuat dengan mengambil 15 jenis *unique value* dari kolom *Area Name*, *Plan*, *Tv Plan*, dan *Advance Promo*. Selain itu, data juga dibuat dengan nilai kolom *Complaint by Customer Service*, *Complaint by Email*, *Complaint by Social Media*, *Complaint by Telegram*, *Complaint by Whatsapp*, dan *Complaint by WIC* masing-masing dengan angka acak dari 1-200.

- b. *Dataset 2*:



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Dataset 2 ini dibuat dengan komposisi sebagai berikut:

- *Dataset* akan memuat data churn sebanyak 1.475 dan data random sebanyak 47.692 atau dengan rasio sekitar 3% : 97%.
- Data random dibuat dengan mengambil semua jenis *unique value* dari kolom *Area Name, Plan, Tv Plan, dan Advance Promo*. Selain itu, data juga dibuat dengan nilai kolom *Complaint by Customer Service, Complaint by Email, Complaint by Social Media, Complaint by Telegram, Complaint by Whatsapp, dan Complaint by WIC* masing-masing dengan angka acak dari 1-200.

c. *Dataset 3:*

Dataset 3 ini dibuat dengan mengambil semua jenis *unique value* dari kolom *Area Name, Plan, Tv Plan, dan Advance Promo*. Selain itu, data juga dibuat dengan nilai kolom *Complaint by Customer Service, Complaint by Email, Complaint by Social Media, Complaint by Telegram, Complaint by Whatsapp, dan Complaint by WIC* masing-masing dengan angka acak dari 1-200. *Dataset 3* ini akan memuat data sebanyak 140.000.

2. Upload *dataset* pada *website* yang telah dibuat
3. Tinjaulah hasil yang diberikan oleh *website* prediksi churn yang telah dibuat
4. Ulangi langkah (2) dan (3) dengan *dataset* yang berbeda

4.1.2.3 Data Hasil Pengujian

Berdasarkan pengujian yang telah dilakukan, data hasil pengujian akan dibagikan menjadi 5 tipe data, yaitu keseluruhan data, data *Area Name*, data *Plan*, data *Tv Plan*, dan data *Advance Promo*.

A. *Dataset 1*

Rincian mengenai *dataset 1* dapat dilihat pada Prosedur Pengujian. Berikut merupakan hasil dari pengujian data *dummy* menggunakan *dataset 1*.

1. Keseluruhan Data



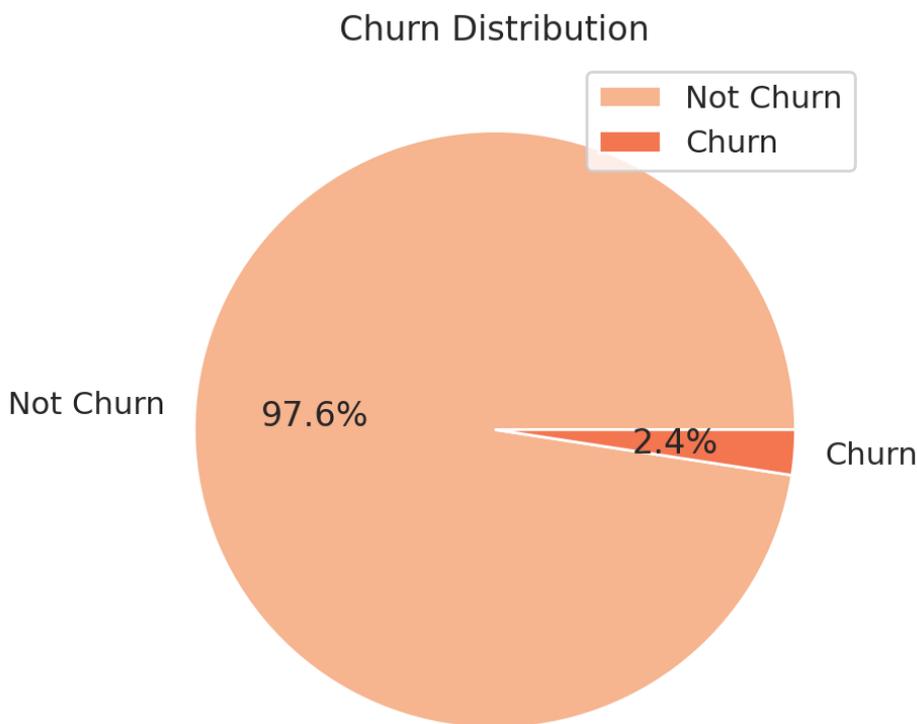
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengunumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Pengelompokkan data ini akan menampilkan hasil prediksi seluruh data, grafik pie distribusi churn, dan hasil distribusi churn seluruh data. Berikut merupakan tampilan dari hasil tersebut.

Area Name	Plan	Tv Plan	Advance Promo	Complaint by Customer Service	Complaint by Email	Complaint by Social Media	Complaint by Telegram	Complaint by Whatsapp	Complaint by WIC	Churn
0 Karawang	BSD Basic 10 Mbps	Combo TV (Jet20)	Advance Payment 6 Month	40	108	169	25	139	145	Not Churn
1 Bandung	BASIC SUPER (Internet Up To 512 Kbps + TV Channel)	Combo TV (Sonic150)	12 Month Advance Payment (Existing Customer)	165	186	46	178	180	109	Not Churn
2 Cilegon	BASIC SUPER (Internet Up To 512 Kbps + TV Channel)	ComboTV 77 Channel	Discount 10000 - PI21 (12Months)	20	161	123	7	129	55	Not Churn
3 Lampung	BSD Pro 300 Mbps	Combo TV (Sonic150)	Credit Card Payment Reward Discount (For Xtra)	64	49	127	127	51	178	Not Churn
4 Ciburur	Basic (promo tv booster)	Cosmic TV	Advance Payment Promo 9+3 (After SA)	43	76	145	100	84	184	Not Churn
5 Jambi	Basic.15+	Combo TV (Fast50)	Discount 15000 - PI21 (11months)	144	28	31	56	66	31	Not Churn
6 Cibubur	BASIC SUPER (Internet Up To 512 Kbps + TV Channel)	Combo TV (Hype75)	Advance Payment Promo 9+3 (New SA)	93	35	62	180	78	47	Not Churn
7 Medan	Basic (promo tv booster)	BASIC	Discount 15000 - PI21 (11months)	18	66	160	117	176	125	Not Churn
8 Karawang	Basic Plus Star	Combo TV (Jet20)	Advance Payment Promo 9+3 (New SA)	113	9	24	84	32	184	Not Churn
9 Bekasi	Basic Plus Star	Cosmic TV	Credit Card Payment Reward Discount (For Xtra)	180	111	20	195	53	1	Not Churn

Gambar 4.6 Hasil Prediksi Seluruh Data



Gambar 4.7 Grafik Pie Distribusi Churn Seluruh Data

Jenis Churn	Jumlah Data	Persentase
0 Churn	1201	2.44%
1 Not Churn	47966	97.56%

Gambar 4.8 Hasil Distribusi Data Churn



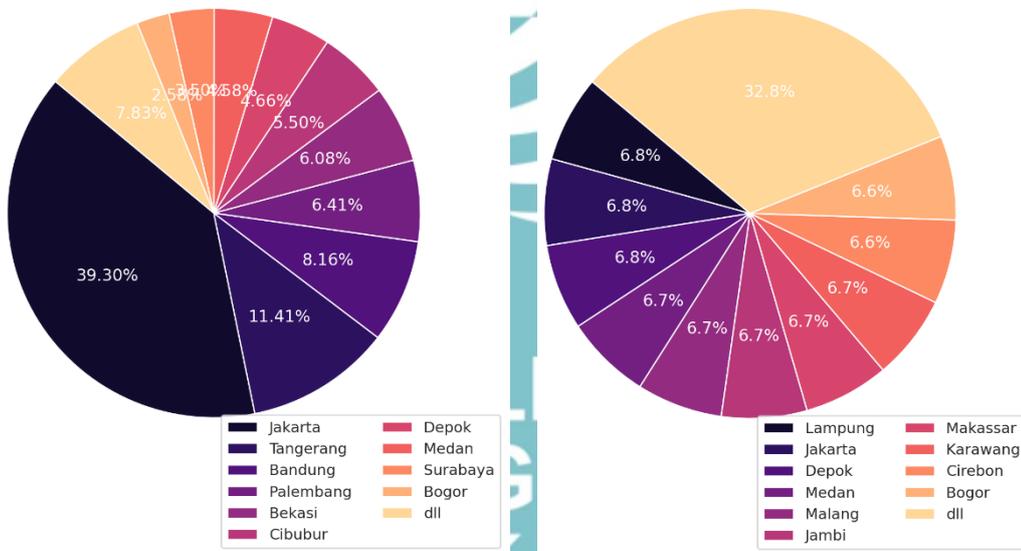
Hak Cipta milik Politeknik Negeri Jakarta

2. Data Area Name

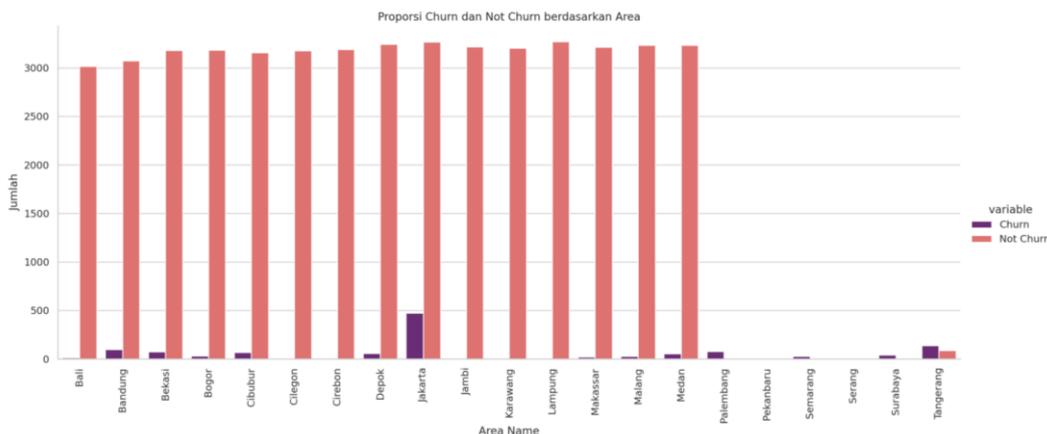
Pengelompokkan data ini akan menampilkan grafik pie distribusi churn, proporsi data yang mengalami churn dan tidak churn Area Name, grafik batang proporsi data yang mengalami churn dan tidak churn Area Name, data jumlah churn dan tidak churn pada masing-masing value pada Area Name berupa tabel, grafik batang jumlah churn dan tidak churn 5 terbanyak pada Area Name, data jumlah churn dan tidak churn 5 terbanyak pada Area Name berupa tabel, grafik jumlah churn dan tidak churn 10 terbanyak pada Area Name, dan data jumlah churn dan tidak churn 10 terbanyak pada Area Name berupa tabel. Berikut merupakan tampilan dari hasil tersebut.

Proportion Count Churned by Area Name

Proportion Count Not Churned by Area Name



Gambar 4.9 Grafik Pie Proporsi Churn & Not Churn Data Area Name



Gambar 4.10 Grafik Bar Proporsi Churn & Not Churn Data Area Name

- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , pennisan karya ilmiah, pennisan laporan, pennisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengunumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



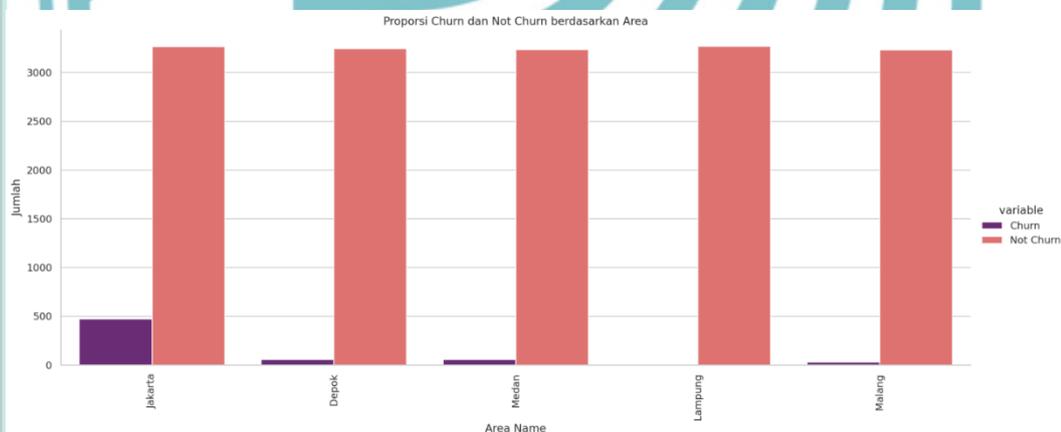
© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

	Area Name	Churn	Not Churn
0	Bali	15	3016
1	Bandung	98	3074
2	Bekasi	73	3179
3	Bogor	31	3182
4	Cibubur	66	3158
5	Cilegon	1	3178
6	Cirebon	0	3189
7	Depok	56	3244
8	Jakarta	472	3267
9	Jambi	0	3218

Gambar 4.11 Tabel Churn & Not Churn Data Area Name



Gambar 4.12 Grafik Bar Top 5 Proporsi Churn & Not Churn Data Area Name

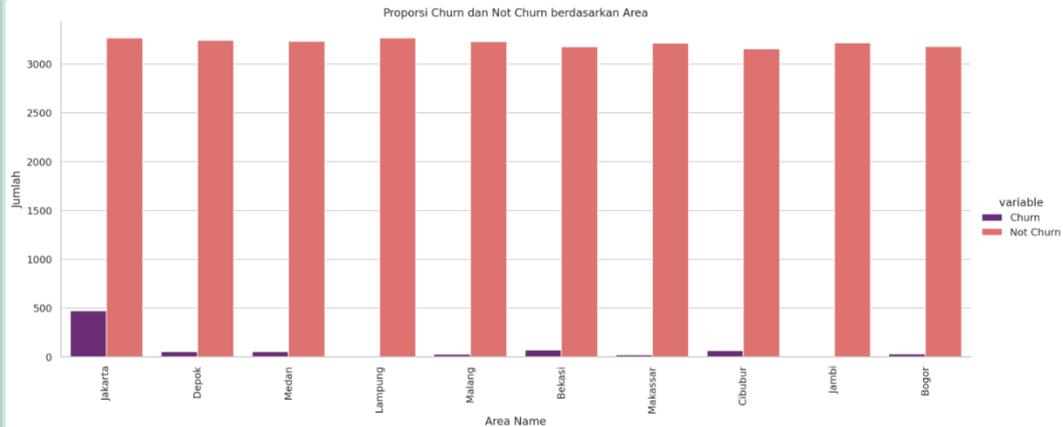
	Area Name	Churn	Not Churn
8	Jakarta	472	3267
7	Depok	56	3244
14	Medan	55	3235
11	Lampung	0	3269
13	Malang	28	3232

Gambar 4.13 Top 5 Proporsi Churn & Not Churn Data Area Name



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penerbitan karya ilmiah, penerbitan laporan, penerbitan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.14 Grafik Bar Top 10 Proporsi Churn & Not Churn Data Area Name

	Area Name	Churn	Not Churn
8	Jakarta	472	3267
7	Depok	56	3244
14	Medan	55	3235
11	Lampung	0	3269
13	Malang	28	3232
2	Bekasi	73	3179
12	Makassar	20	3215
4	Cibubur	66	3158
9	Jambi	0	3218
3	Bogor	31	3182

Gambar 4.15 Tabel Top 10 Proporsi Churn & Not Churn Data Area Name

	Area Name	Count Churned	Persentase Dari Data Churn
0	Bekasi	73	6.08%
1	Cibubur	66	5.50%
2	Surabaya	42	3.50%
3	Tangerang	137	11.41%
4	Depok	56	4.66%
5	Medan	55	4.58%
6	Palembang	77	6.41%
7	Jakarta	472	39.30%
8	Semarang	26	2.16%
9	Bogor	31	2.58%

Gambar 4.16 Tabel Churn & Persentase Data – Area Name



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

	Area Name	Count Not Churned	Persentase Dari Data Not Churn
0	Karawang	3202	6.68%
1	Bandung	3074	6.41%
2	Cilegon	3178	6.63%
3	Lampung	3269	6.82%
4	Cibubur	3158	6.58%
5	Jambi	3218	6.71%
6	Medan	3235	6.74%
7	Bekasi	3179	6.63%
8	Depok	3244	6.76%
9	Malang	3232	6.74%

Gambar 4.17 Tabel Not Churn & Persentase Data – Area Name

3. Data Plan

Pengelompokkan data ini akan menampilkan grafik pie distribusi churn, proporsi data yang mengalami churn dan tidak churn *Plan*, grafik batang proporsi data yang mengalami churn dan tidak churn *Plan*, data jumlah churn dan tidak churn pada masing-masing *value* pada *Plan* berupa tabel, grafik batang jumlah churn dan tidak churn 5 terbanyak pada *Plan*, data jumlah churn dan tidak churn 5 terbanyak pada *Plan* berupa tabel, grafik jumlah churn dan tidak churn 10 terbanyak pada *Plan*, dan data jumlah churn dan tidak churn 10 terbanyak pada *Plan* berupa tabel. Berikut merupakan tampilan dari hasil tersebut.

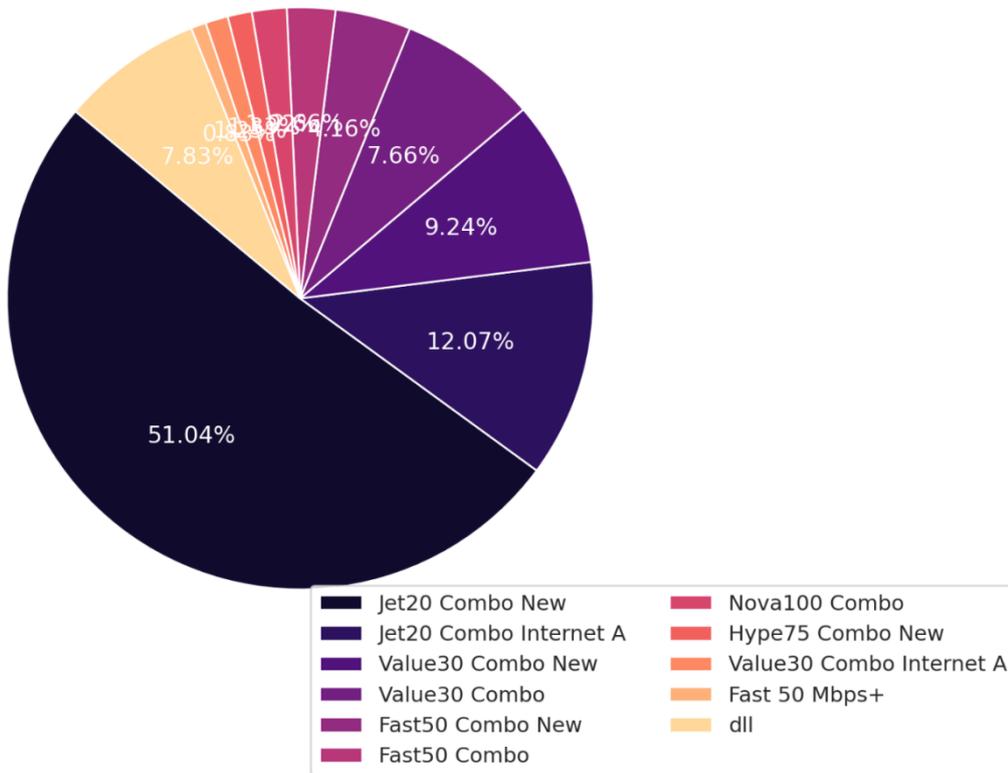


Hak Cipta milik Politeknik Negeri Jakarta

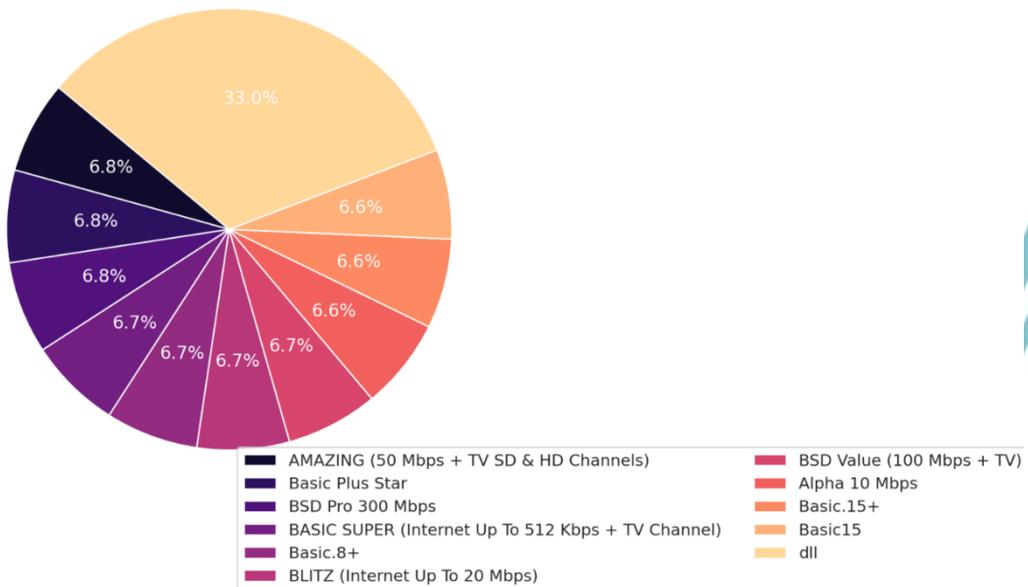
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Proportion Count Churned by Plan



Proportion Count Not Churned by Plan



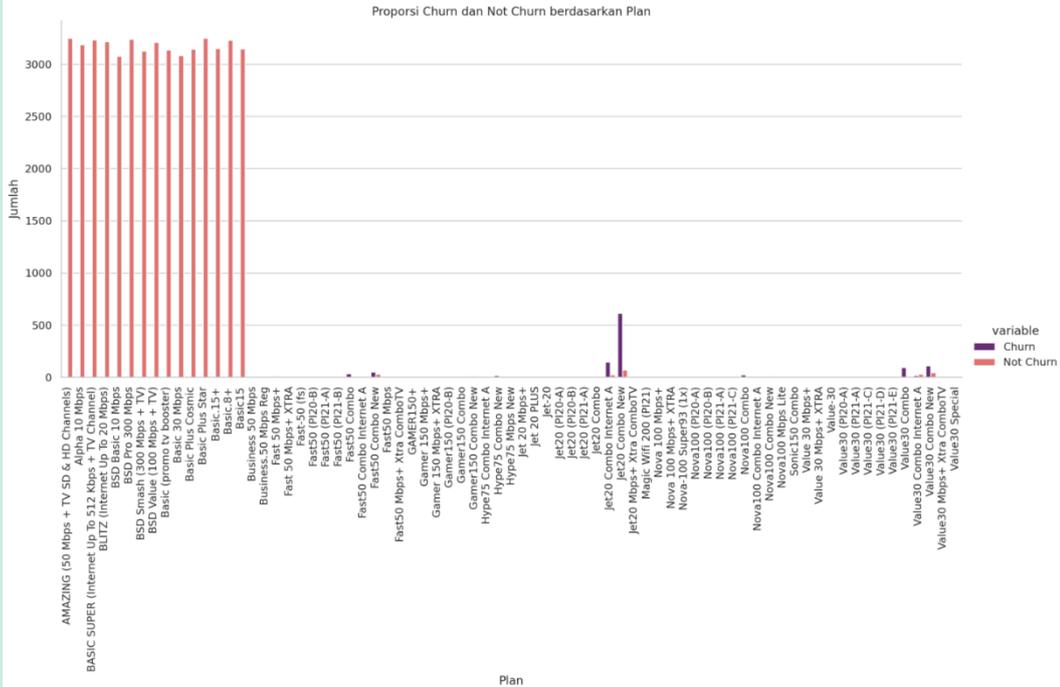
Gambar 4.18 Grafik Pie Proporsi Churn & Not Churn Data Plan



Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.19 Grafik Bar Proporsi Churn & Not Churn Data Plan

Plan	Churn	Not Churn
0 AMAZING (50 Mbps + TV SD & HD Channels)	0	3250
1 Alpha 10 Mbps	0	3187
2 BASIC SUPER (Internet Up To 512 Kbps + TV Channel)	0	3235
3 BLITZ (Internet Up To 20 Mbps)	0	3218
4 BSD Basic 10 Mbps	0	3076
5 BSD Pro 300 Mbps	0	3242
6 BSD Smash (300 Mbps + TV)	0	3128
7 BSD Value (100 Mbps + TV)	0	3211
8 Basic (promo tv booster)	0	3138
9 Basic 30 Mbps	0	3084

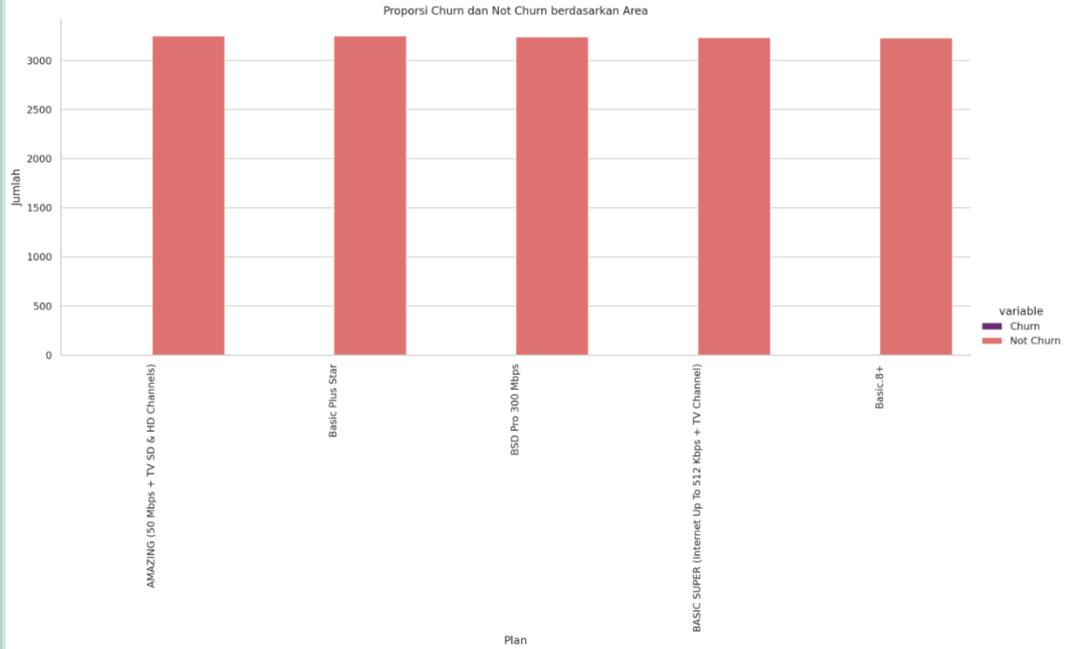
Gambar 4.20 Tabel Churn & Not Churn Data Plan



Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.21 Grafik Bar Top 5 Proporsi Churn & Not Churn Data *Plan*

	Plan	Churn	Not Churn
0	AMAZING (50 Mbps + TV SD & HD Channels)	0	3250
11	Basic Plus Star	0	3250
5	BSD Pro 300 Mbps	0	3242
2	BASIC SUPER (Internet Up To 512 Kbps + TV Channel)	0	3235
13	Basic.8+	0	3229

Gambar 4.22 Tabel Top 5 Proporsi Churn & Not Churn Data *Plan*

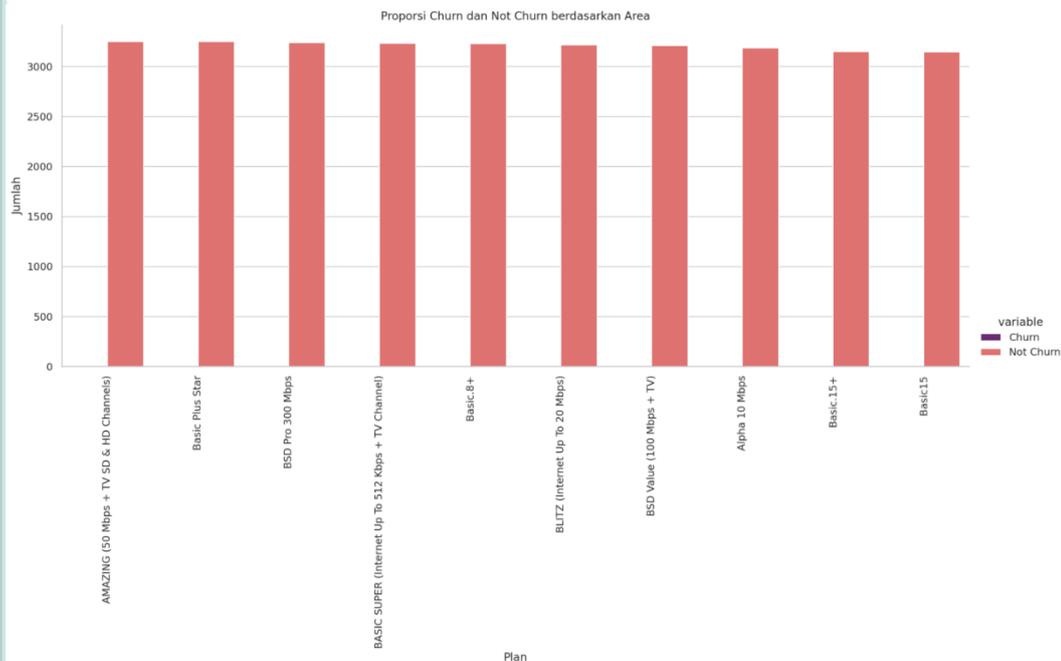




Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.23 Grafik Bar Top 10 Proporsi Churn & Not Churn Data *Plan*

	Plan	Churn	Not Churn
0	AMAZING (50 Mbps + TV SD & HD Channels)	0	3250
11	Basic Plus Star	0	3250
5	BSD Pro 300 Mbps	0	3242
2	BASIC SUPER (Internet Up To 512 Kbps + TV Channel)	0	3235
13	Basic.8+	0	3229
3	BLITZ (Internet Up To 20 Mbps)	0	3218
7	BSD Value (100 Mbps + TV)	0	3211
1	Alpha 10 Mbps	0	3187
12	Basic.15+	0	3152
14	Basic15	0	3148

Gambar 4.24 Tabel Top 10 Proporsi Churn & Not Churn Data *Plan*



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

	Plan	Count Churned	Persentase Dari Data Churn
0	Value30 (PI20-A)	2	0.17%
1	Fast50 (PI21-A)	2	0.17%
2	Value30 (PI21-A)	4	0.33%
3	Fast50 Combo New	50	4.16%
4	Hype75 Combo New	16	1.33%
5	Gamer150 (PI20-B)	1	0.08%
6	Nova100 (PI20-A)	1	0.08%
7	Jet20 (PI20-B)	1	0.08%
8	Nova100 (PI20-B)	1	0.08%
9	Fast50 (PI21-B)	1	0.08%

Gambar 4.25 Tabel Churn & Persentase Data – Plan

	Area Name	Count Not Churned	Persentase Dari Data Not Churn
0	Karawang	3202	6.68%
1	Bandung	3074	6.41%
2	Cilegon	3178	6.63%
3	Lampung	3269	6.82%
4	Cibubur	3158	6.58%
5	Jambi	3218	6.71%
6	Medan	3235	6.74%
7	Bekasi	3179	6.63%
8	Depok	3244	6.76%
9	Malang	3232	6.74%

Gambar 4.26 Tabel Not Churn & Persentase Data – Plan

4. Data Tv Plan

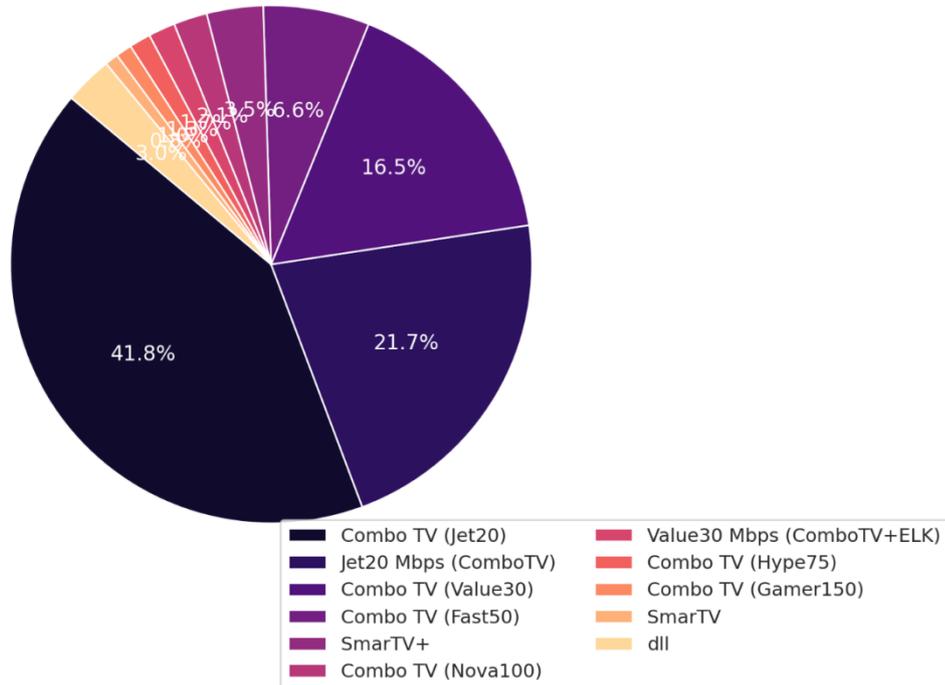
Pengelompokkan data ini akan menampilkan grafik pie distribusi churn, proporsi data yang mengalami churn dan tidak churn *Tv Plan*, grafik batang proporsi data yang mengalami churn dan tidak churn *Tv Plan*, data jumlah churn dan tidak churn pada masing-masing *value* pada *Tv Plan* berupa tabel, grafik batang jumlah churn dan tidak churn 5 terbanyak pada *Tv Plan*, data jumlah churn dan tidak churn 5 terbanyak pada *Tv Plan* berupa tabel, grafik jumlah churn dan tidak churn 10 terbanyak pada *Tv Plan*, dan data jumlah churn dan tidak churn 10



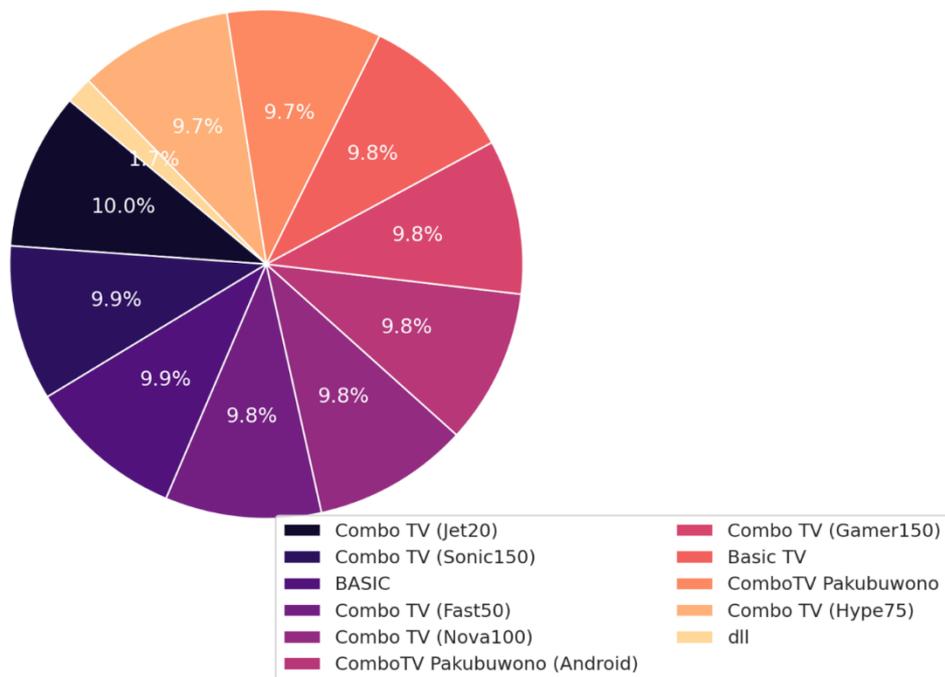
© Hak Cipta milik Politeknik Negeri Jakarta

terbanyak pada *Tv Plan* berupa tabel. Berikut merupakan tampilan dari hasil tersebut.

Proportion Count Churned by Tv Plan



Proportion Count Not Churned by Tv Plan



Gambar 4.27 Grafik Pie Proporsi Churn & Not Churn Data *Tv Plan*

Hak Cipta :

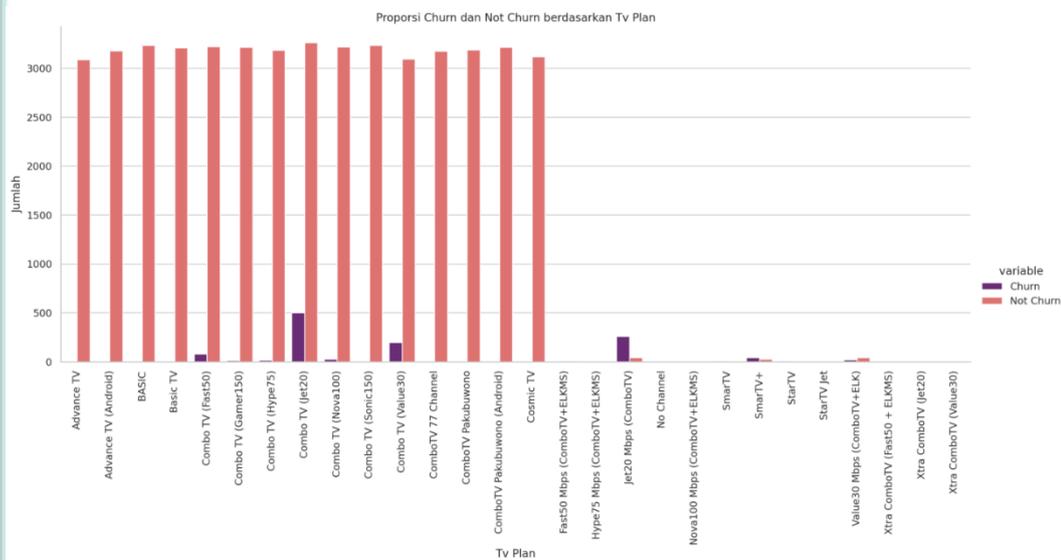
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.28 Grafik Bar Proporsi Churn & Not Churn Data *Tv Plan*

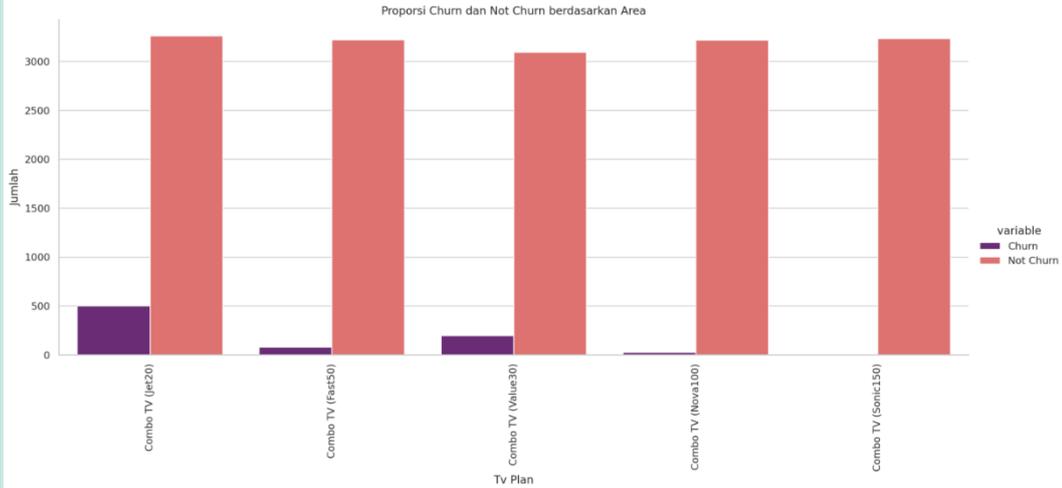
	Tv Plan	Churn	Not Churn
0	Advance TV	2	3088
1	Advance TV (Android)	0	3176
2	BASIC	0	3235
3	Basic TV	7	3206
4	Combo TV (Fast50)	79	3221
5	Combo TV (Gamer150)	12	3214
6	Combo TV (Hype75)	16	3183
7	Combo TV (Jet20)	502	3261
8	Combo TV (Nova100)	25	3219
9	Combo TV (Sonic150)	3	3235

Gambar 4.29 Tabel Churn & Not Churn Data *Tv Plan*



Hak Cipta :

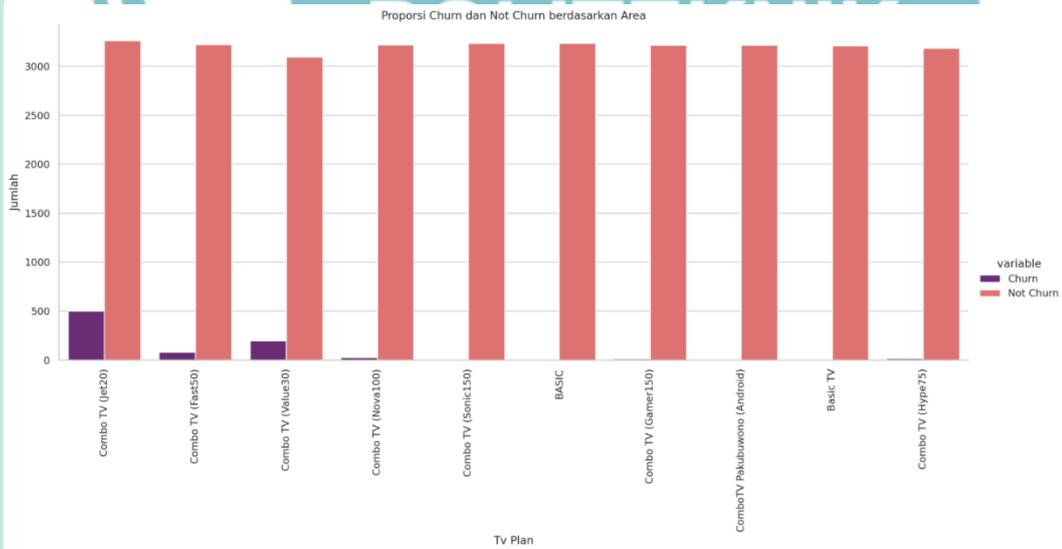
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.30 Grafik Bar Top 5 Proporsi Churn & Not Churn Data Tv Plan

	Tv Plan	Churn	Not Churn
7	Combo TV (Jet20)	502	3261
4	Combo TV (Fast50)	79	3221
10	Combo TV (Value30)	198	3095
8	Combo TV (Nova100)	25	3219
9	Combo TV (Sonic150)	3	3235

Gambar 4.31 Tabel Top 5 Proporsi Churn & Not Churn Data Tv Plan



Gambar 4.32 Grafik Bar Top 10 Proporsi Churn & Not Churn Data Area Name



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

	Tv Plan	Churn	Not Churn
7	Combo TV (Jet20)	502	3261
4	Combo TV (Fast50)	79	3221
10	Combo TV (Value30)	198	3095
8	Combo TV (Nova100)	25	3219
9	Combo TV (Sonic150)	3	3235
2	BASIC	0	3235
5	Combo TV (Gamer150)	12	3214
13	ComboTV Pakubuwo (Android)	0	3216
3	Basic TV	7	3206
6	Combo TV (Hype75)	16	3183

Gambar 4.33 Tabel Top 10 Proporsi Churn & Not Churn Data *Tv Plan*

	Tv Plan	Count Churned	Persentase Dari Data Churn
0	StarTV	9	0.75%
1	SmarTV	10	0.83%
2	Basic TV	7	0.58%
3	Combo TV (Fast50)	79	6.58%
4	Combo TV (Hype75)	16	1.33%
5	SmarTV+	42	3.50%
6	Jet20 Mbps (ComboTV)	261	21.73%
7	StarTV Jet	5	0.42%
8	No Channel	1	0.08%
9	Xtra ComboTV (Value30)	1	0.08%

Gambar 4.34 Tabel Churn & Persentase Data – *Tv Plan*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

	Tv Plan	Count Not Churned	Persentase Dari Data Not Churn
0	Combo TV (Jet20)	3261	6.80%
1	Combo TV (Sonic150)	3235	6.74%
2	ComboTV 77 Channel	3174	6.62%
3	Cosmic TV	3119	6.50%
4	Combo TV (Fast50)	3221	6.72%
5	Combo TV (Hype75)	3183	6.64%
6	BASIC	3235	6.74%
7	Combo TV (Value30)	3095	6.45%
8	Combo TV (Gamer150)	3214	6.70%
9	Advance TV (Android)	3176	6.62%

Gambar 4.35 Tabel Not Churn & Persentase Data – *Tv Plan*

5. *Data Advance Promo*

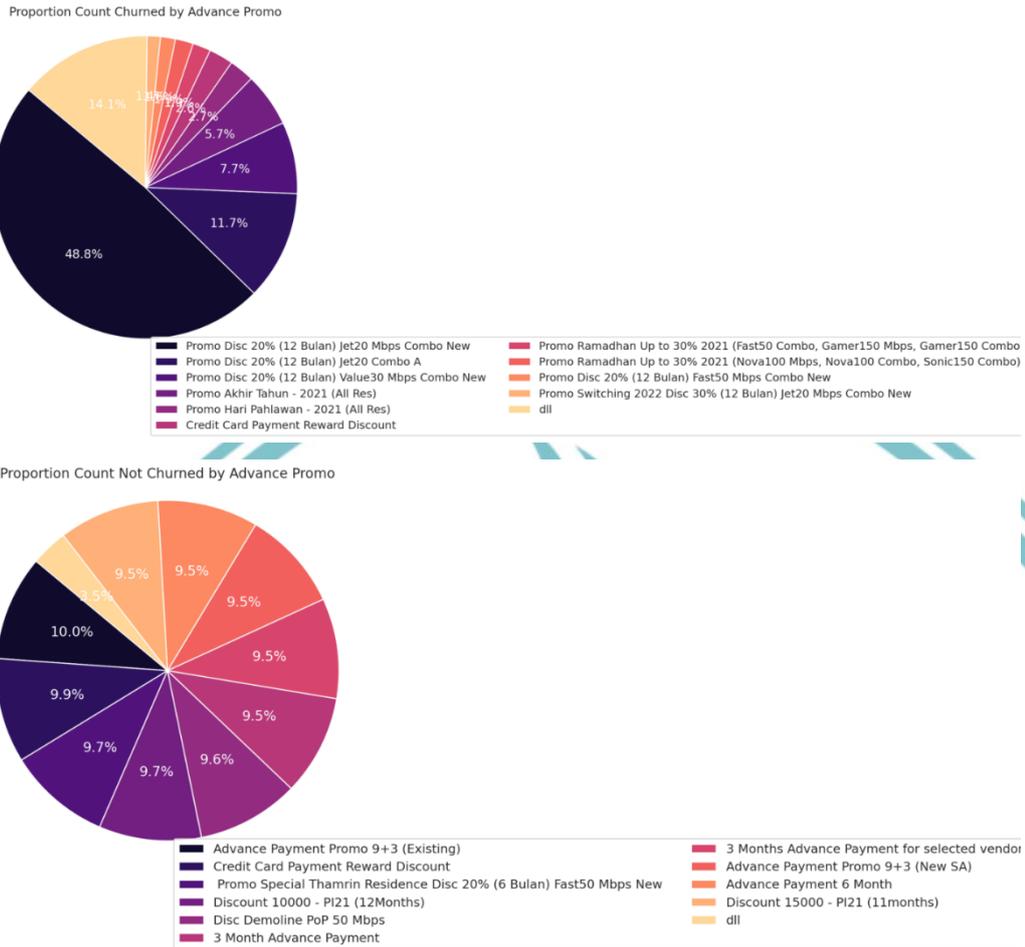
Pengelompokkan data ini akan menampilkan grafik pie distribusi churn, proporsi data yang mengalami churn dan tidak churn *Advance Promo*, grafik batang proporsi data yang mengalami churn dan tidak churn *Advance Promo*, data jumlah churn dan tidak churn pada masing-masing *value* pada *Advance Promo* berupa tabel, grafik batang jumlah churn dan tidak churn 5 terbanyak pada *Advance Promo*, data jumlah churn dan tidak churn 5 terbanyak pada *Advance Promo* berupa tabel, grafik jumlah churn dan tidak churn 10 terbanyak pada *Advance Promo*, dan data jumlah churn dan tidak churn 10 terbanyak pada *Advance Promo* berupa tabel. Berikut merupakan tampilan dari hasil tersebut.



Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.36 Grafik Bar Proporsi Churn & Not Churn Data Advance Promo

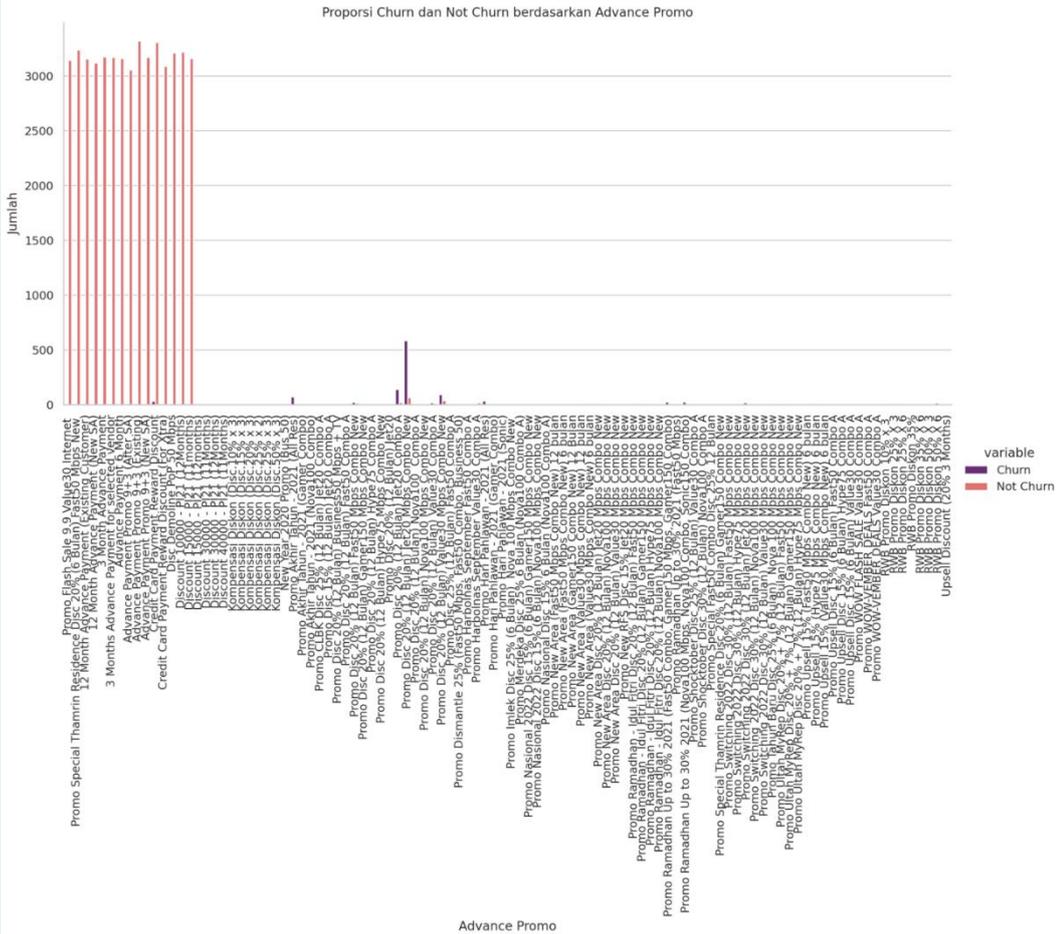




Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.37 Grafik Bar Proporsi Churn & Not Churn Data Advance Promo

Advance Promo	Churn	Not Churn
0 Promo Flash Sale 9.9 Value30 Internet	0	3145
1 Promo Special Thamrin Residence Disc 20% (6 Bulan) Fast50 Mbps New	0	3240
2 12 Month Advance Payment (Existing Customer)	0	3159
3 12 Month Advance Payment (New SA)	0	3122
4 3 Month Advance Payment	0	3176
5 3 Months Advance Payment for selected vendor	0	3172
6 Advance Payment 6 Month	0	3162
7 Advance Payment Promo 9+3 (After SA)	0	3057
8 Advance Payment Promo 9+3 (Existing)	3	3321
9 Advance Payment Promo 9+3 (New SA)	0	3171

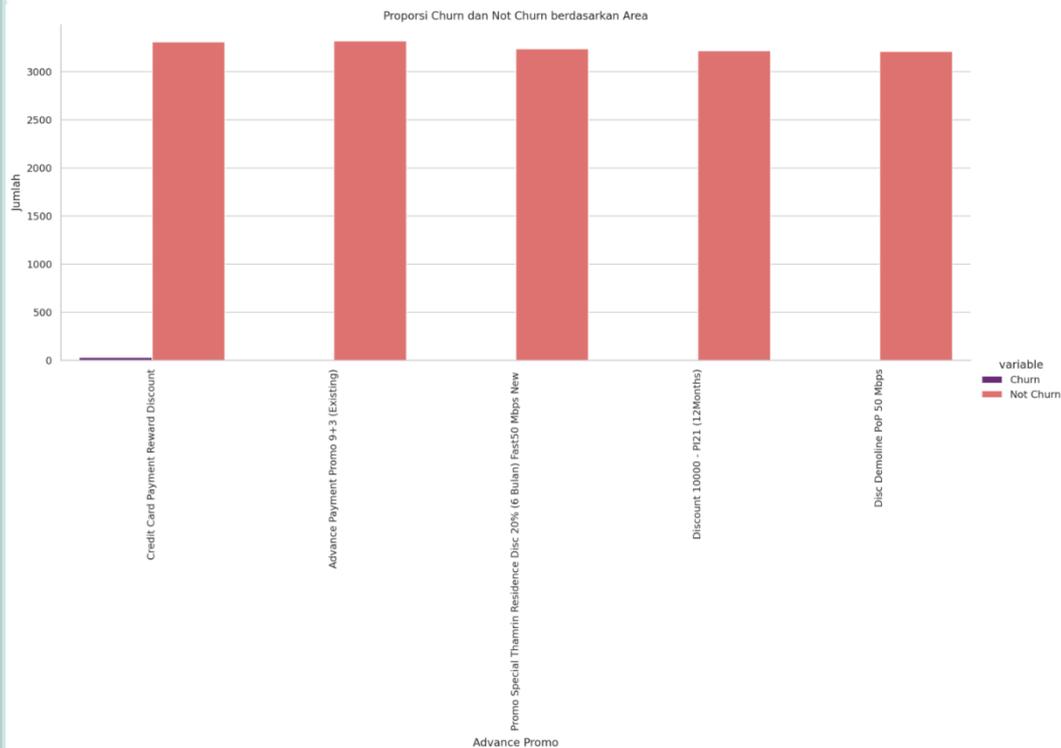
Gambar 4.38 Tabel Churn & Not Churn Data Advance Promo



Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.39 Grafik Bar Top 5 Proporsi Churn & Not Churn Data *Advance Promo*

Advance Promo	Churn	Not Churn
10 Credit Card Payment Reward Discount	31	3309
8 Advance Payment Promo 9+3 (Existing)	3	3321
1 Promo Special Thamrin Residence Disc 20% (6 Bulan) Fast50 Mbps New	0	3240
13 Discount 10000 - PI21 (12Months)	0	3219
12 Disc Demoline PoP 50 Mbps	0	3211

Gambar 4.40 Tabel Top 5 Proporsi Churn & Not Churn Data *Advance Promo*



Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.41 Grafik Bar Top 10 Proporsi Churn & Not Churn Data *Advance Promo*

Advance Promo	Churn	Not Churn
10 Credit Card Payment Reward Discount	31	3309
8 Advance Payment Promo 9+3 (Existing)	3	3321
1 Promo Special Thamrin Residence Disc 20% (6 Bulan) Fast50 Mbps New	0	3240
13 Discount 10000 - PI21 (12Months)	0	3219
12 Disc Demoline PoP 50 Mbps	0	3211
4 3 Month Advance Payment	0	3176
5 3 Months Advance Payment for selected vendor	0	3172
9 Advance Payment Promo 9+3 (New SA)	0	3171
14 Discount 15000 - PI21 (11months)	2	3161
6 Advance Payment 6 Month	0	3162

Gambar 4.42 Tabel Top 10 Proporsi Churn & Not Churn Data *Advance Promo*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Advance Promo	Count Churned	Persentase Dari Data Churn
0 Kompensasi Diskon (Disc.25% x 3)	2	0.17%
1 RWB Promo Diskon 50% X 6	12	1.00%
2 Discount 30000 - PI21 (11Months)	1	0.08%
3 Discount 15000 - PI21 (11months)	2	0.17%
4 Promo Upsell 15% (Fast50 Mbps Combo New) 6 bulan	6	0.50%
5 Promo Upsell 15% (Hype75 Mbps Combo New) 6 bulan	8	0.67%
6 RWB Promo Diskon 25% X 3	11	0.92%
7 RWB Promo Diskon 35% X 6	7	0.58%
8 Promo Disc 20% (12 Bulan) Jet20 Mbps Combo New	586	48.79%
9 Advance Payment Promo 9+3 (Existing)	3	0.25%

Gambar 4.43 Tabel Churn & Persentase Data – Advance Promo

Advance Promo	Count Not Churned	Persentase Dari Data Not Churn
0 Advance Payment 6 Month	3162	6.59%
1 12 Month Advance Payment (Existing Customer)	3159	6.59%
2 Discount 10000 - PI21 (12Months)	3219	6.71%
3 Credit Card Payment Reward Discount (For Xtra)	3091	6.44%
4 Advance Payment Promo 9+3 (After SA)	3057	6.37%
5 Discount 15000 - PI21 (11months)	3161	6.59%
6 Advance Payment Promo 9+3 (New SA)	3171	6.61%
7 Credit Card Payment Reward Discount	3309	6.90%
8 Advance Payment Promo 9+3 (Existing)	3321	6.92%
9 3 Months Advance Payment for selected vendor	3172	6.61%

Gambar 4.44 Tabel Not Churn & Persentase Data – Advance Promo

B. *Dataset 2 dan 3*

Hasil pengujian untuk *dataset 2* dan *dataset 3* akan ditampilkan pada L-9 dan L-10.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

4.1.2.4 Analisa Hasil Pengujian

A. *Dataset 1*

Dataset 1 terdiri atas beberapa jenis pengelompokkan, antara lain:

1. Keseluruhan Data

Gambar 4.6 dapat ditampilkan dengan baik hal ini menandakan bahwa proses prediksi yang dilakukan oleh model *machine learning* berjalan dengan baik. Hasil yang ditampilkan pada Gambar 4.7 merupakan distribusi data yang mengalami churn dan tidak mengalami churn. Terlihat bahwa dari total data sebanyak 49.167, terdapat 97,6% data yang tidak mengalami churn dan sebanyak 2,4% data yang mengalami churn. Gambar 4.8 menunjukkan hasil jumlah data yang mengalami churn sebanyak 1.201 data dengan nilai persentase sebesar 2,44%. Sementara itu, data yang tidak mengalami churn sebanyak 47.966 data dengan nilai persentase sebesar 97,56%. Meninjau dari *dataset* yang dibuat, *dataset* yang melalui prediksi memuat data churn sebanyak 1.475. Jika dibandingkan dengan data churn dari *dataset A* yang telah disiapkan dengan hasil dari prediksi yang dikeluarkan oleh *website* prediksi churn, terlihat bahwa terdapat data yang mengalami salah prediksi. Rasio dari data churn kedua data sebesar $\pm 8,1:10$, dengan angka 0,8142372881 untuk *output* dari prediksi churn. Dengan ini dapat diketahui tingkat *error* dari prediksi ini berada di angka 0,1857627119.

2. *Area Name*

Berdasarkan Gambar 4.17 dapat ditarik kesimpulan data sebagai berikut.

Tabel 4.2 Data Rank *Area Name*

Data <i>Area Name</i>		
Churn / Not Churn	Nama Area	Rank
	Jakarta	1
	Tangerang	2
Churn	Bandung	3
	Palembang	4
	Bekasi	5



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Cibubur	6
Depok	7
Medan	8
Surabaya	9
Bogor	10
<hr/>	
Lampung	1
Jakarta	2
Depok	3
Medan	4
Malang	5
Jambi	6
Makassar	7
Karawang	8
Cirebon	9
Bogor	10

Not Churn

Terlihat untuk *Area Name* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Area Name* Jakarta. Sementara itu, *Area Name* yang tidak mengalami churn paling banyak yaitu berada pada *Area Name* Lampung.

Gambar 4.9 menampilkan grafik batang proporsi data churn & data tidak churn pada *Area Name*. Pada gambar tersebut, terlihat bahwa grafik batang *Area Name* “Jakarta” pada variabel churn sangatlah mendominasi dibandingkan dengan *Area Name* lainnya. Gambar 4.10 menampilkan 10 data *Area Name*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*. Pada Gambar 4.9 di dapat terlihat bahwa data sulit untuk dibaca. Oleh karena itu, diperlukannya penampilan data dengan sistem *ranking*. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Di sini akan ditampilkan data “Top 5” dan “Top 10” data *Area Name*. Data tersebut ditampilkan pada Gambar 4.11 dan Gambar 4.13.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Berdasarkan Gambar 4.11 dan Gambar 4.12, dapat ditarik kesimpulan bahwa data yang memiliki jumlah data gabungan antara Churn dan Not Churn 5 teratas antara lain, Jakarta, Depok, Medan, Lampung, dan Malang. Jakarta dengan total Churn & Not Churn 3739 data, Depok dengan total Churn & Not Churn 3300 data, Medan dengan total Churn & Not Churn 3290 data, Lampung dengan total Churn & Not Churn 3269 data, dan Malang dengan total Churn & Not Churn 3260 data. Berdasarkan Gambar 4.13 dan Gambar 4.14, dapat ditarik kesimpulan bahwa data yang memiliki jumlah data gabungan antara Churn dan Not Churn 10 teratas antara lain, Jakarta, Depok, Medan, Lampung, Malang, Bekasi, Makassar, Cibubur, Jambi, dan Bogor. Jakarta dengan total Churn & Not Churn 3739 data, Depok dengan total Churn & Not Churn 3300 data, Medan dengan total Churn & Not Churn 3290 data, Lampung dengan total Churn & Not Churn 3269 data, Malang dengan total Churn & Not Churn 3260 data, Bekasi dengan total Churn & Not Churn 3252 data, Makassar dengan total Churn & Not Churn 3235 data, Cibubur dengan total Churn & Not Churn 3224 data, Jambi dengan total Churn & Not Churn 3218 data, dan Bogor dengan total Churn & Not Churn 3213 data.

Gambar 4.15 menampilkan data *Area Name* yang mengalami churn. Pada data tersebut ditampilkan jumlah data yang mengalami churn dan persentase dari data churn untuk *Area Name* keseluruhan. Sementara itu, Gambar 4.16 menampilkan data *Area Name* yang tidak mengalami churn. Pada data tersebut ditampilkan jumlah data yang tidak mengalami churn dan persentase dari data *not churn* untuk *Area Name* keseluruhan. Data yang ditampilkan dibatasi hanya 10 baris, agar tampilan pada *website* lebih tertata dan tidak membuang-buang ruang pada *website*.

3. *Plan*

Dari Gambar 4.17 di atas dapat ditarik kesimpulan data sebagai berikut.

Tabel 4.3 Data Rank *Plan*

Data <i>Plan</i>		
Churn / Not Churn	Nama <i>Plan</i>	Ran k



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Churn	Jet20 Combo New	1
	Jet20 Combo Internet A	2
	Value30 Combo New	3
	Value30 Combo	4
	Fast50 Combo New	5
	Fast50 Combo	6
	Nova100 Combo	7
	Hype75 Combo New	8
	Value30 Combo Internet A	9
	Fast 50 Mbps+	10
Not Churn	AMAZING (50 Mbps + Tv SD & HD Channels)	1
	Basic Plus Star	2
	BSD Pro 300 Mbps	3
	BASIC SUPER (Internet Up To 521 Kbps + Tv Channel)	4
	Basic.8+	5
	BLITZ (Internet Up To 20 Mbps)	6
	BSD Value (100 Mbps + Tv)	7
	Alpha 10 Mbps	8
	Basic.15+	9
	Basic15	10

Terlihat untuk *Plan* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Plan* “Jet20 Combo New”. Sementara itu, *Plan* yang tidak mengalami churn paling banyak yaitu berada pada *Plan* “AMAZING (50 Mbps + Tv SD & HD Channels)”.

Gambar 4.18 menampilkan grafik batang proporsi data churn & data tidak churn pada *Plan*. Pada gambar tersebut, terlihat bahwa grafik batang *Plan* “Jet20 Combo New” pada variabel churn sangatlah mendominasi dibandingkan dengan *Plan* lainnya. Gambar 4.19 menampilkan 10 data *Plan*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak



© Hak Cipta milik Politeknik Negeri Jakarta

membuang—buang ruang pada *website*. Pada Gambar 4.18 di dapat terlihat bahwa data sulit untuk dibaca. Oleh karena itu, diperlukannya penampilan data dengan sistem *ranking*. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Di sini akan ditampilkan data “Top 5” dan “Top 10” data *Plan*. Data tersebut ditampilkan pada Gambar 4.20 dan Gambar 4.22.

Berdasarkan Gambar 4.20 dan Gambar 4.21, dapat ditarik kesimpulan bahwa data yang memiliki jumlah data gabungan antara Churn dan Not Churn 5 teratas antara lain, “AMAZING (50 Mbps + Tv SD & HD Channels)”, “Basic Plus Star”, “BSD Pro 300 Mbps”, “BASIC SUPER (Internet Up To 512 Kbps + Tv Channel)”, “Basic.8+”. “AMAZING (50 Mbps + Tv SD & HD Channels)” dengan total Churn & Not Churn 3250 data, “Basic Plus Star” dengan total Churn & Not Churn 3250 data, “BSD Pro 300 Mbps” dengan total Churn & Not Churn 3242 data, “BASIC SUPER (Internet Up To 512 Kbps + Tv Channel)” dengan total Churn & Not Churn 3235 data, dan “Basic.8+” dengan total Churn & Not Churn 3229 data. Berdasarkan Gambar 4.22 dan Gambar 4.23, dapat ditarik kesimpulan bahwa data yang memiliki jumlah data gabungan antara Churn dan Not Churn 5 teratas antara lain, “AMAZING (50 Mbps + Tv SD & HD Channels)”, “Basic Plus Star”, “BSD Pro 300 Mbps”, “BASIC SUPER (Internet Up To 512 Kbps + Tv Channel)”, “Basic.8+”, “BLITZ (Internet Up To 20 Mbps)”, “BSD Value (100 Mbps + Tv)”, “Alpha 10 Mbps”, “Basic.15+”, “Basic15”. “AMAZING (50 Mbps + Tv SD & HD Channels)” dengan total Churn & Not Churn 3250 data, “Basic Plus Star” dengan total Churn & Not Churn 3250 data, “BSD Pro 300 Mbps” dengan total Churn & Not Churn 3242 data, “BASIC SUPER (Internet Up To 512 Kbps + Tv Channel)” dengan total Churn & Not Churn 3235 data, dan “Basic.8+” dengan total Churn & Not Churn 3229 data, “BLITZ (Internet Up To 20 Mbps)” dengan total Churn & Not Churn 3218 data, “BSD Value (100 Mbps + Tv)” dengan total Churn & Not Churn 3211 data, “Alpha 10 Mbps” dengan total Churn & Not Churn 3187 data, “Basic.15+” dengan total Churn & Not Churn 3152 data, “Basic15” dengan total Churn & Not Churn 3148 data.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Gambar 4.24 menampilkan data *Plan* yang mengalami churn. Pada data tersebut ditampilkan jumlah data yang mengalami churn dan persentase dari data churn untuk *Plan* keseluruhan. Sementara itu, Gambar 4.25 menampilkan data *Plan* yang tidak mengalami churn. Pada data tersebut ditampilkan jumlah data yang tidak mengalami churn dan persentase dari data *not* churn untuk *Plan* keseluruhan. Data yang ditampilkan dibatasi hanya 10 baris, agar tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*.

4. *Tv Plan*

Dari Gambar 4.11 dapat ditarik kesimpulan data sebagai berikut.

Tabel 4.4 Data Rank *Tv Plan*

Data <i>Tv Plan</i>		
Churn / Not Churn	Nama <i>Plan</i>	Rank
Churn	Combo <i>Tv</i> (Jet20)	1
	Jet20 Mbps (Combo <i>Tv</i>)	2
	Combo <i>Tv</i> (Value30)	3
	Combo <i>Tv</i> (Fast50)	4
	Smar <i>Tv</i> +	5
	Combo <i>Tv</i> (Nova100)	6
	Value30 Mbps (Combo <i>Tv</i> +ELK)	7
	Combo <i>Tv</i> (Hype75)	8
	Combo <i>Tv</i> (Gamer150)	9
	Smar <i>Tv</i>	10
Not Churn	Combo <i>Tv</i> (jet20)	1
	BASIC	2
	Combo <i>Tv</i> (Sonic150)	3
	Combo <i>Tv</i> (Fast50)	4
	Combo <i>Tv</i> (Nova100)	5
	Combo <i>Tv</i> Pakubuwono (Android)	6
	Combo <i>Tv</i> (Gamer150)	7
	Basic <i>Tv</i>	8

ComboTv Pakubuwono	9
Combo Tv (Hype75)	10

Terlihat untuk *Tv Plan* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Tv Plan* “Combo Tv (Jet20)”. Sementara itu, *Tv Plan* yang tidak mengalami churn paling banyak yaitu berada pada *Tv Plan* “ComboTv (jet20)”.

Gambar 4.27 menampilkan grafik batang proporsi data churn & data tidak churn pada *Tv Plan*. Pada gambar tersebut, terlihat bahwa grafik batang *Tv Plan* “Combo Tv (Jet20)” pada variabel churn sangatlah mendominasi dibandingkan dengan *Plan* lainnya. Gambar 4.28 menampilkan 10 data *Tv Plan*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*. Pada Gambar 4.27 di dapat terlihat bahwa data sulit untuk dibaca. Oleh karena itu, diperlukannya penampilan data dengan sistem *ranking*. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Di sini akan ditampilkan data “Top 5” dan “Top 10” data *Tv Plan*. Data tersebut ditampilkan pada Gambar 4.29 dan Gambar 4.31.

Berdasarkan Gambar 4.29 dan Gambar 4.30, dapat ditarik kesimpulan bahwa data yang memiliki jumlah data gabungan antara Churn dan Not Churn 5 teratas antara lain, “Combo Tv (Jet20)”, “Combo Tv (Fast50)”, “Combo Tv (Value30)”, “Combo Tv (Nova100)”, dan “Combo Tv(Sonic150)”. “Combo Tv (Jet20)” dengan total Churn & Not Churn 3763 data, “Combo Tv (Fast50)” dengan total Churn & Not Churn 3300 data, “Combo Tv (Value30)” dengan total Churn & Not Churn 3293 data, “Combo Tv (Nova100)” dengan total Churn & Not Churn 3244 data, dan “Combo Tv (Sonic150)” dengan total Churn & Not Churn 3238 data. Berdasarkan Gambar 4.31 dan Gambar 4.32, dapat ditarik kesimpulan bahwa data yang memiliki jumlah data gabungan antara Churn dan Not Churn 10 teratas antara lain, “Combo Tv (Jet20)”, “Combo Tv (Fast50)”, “Combo Tv (Value30)”, “Combo Tv (Nova100)”, “Combo Tv(Sonic150)”, “BASIC”, “Combo Tv (Gamer150)”,



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

“Combo Tv Pakubuwono (Android)”, “Basic Tv”, dan “Combo Tv (Hype75)”. “Combo Tv (Jet20)” dengan total Churn & Not Churn 3763 data, “Combo Tv (Fast50)” dengan total Churn & Not Churn 3300 data, “Combo Tv (Value30)” dengan total Churn & Not Churn 3293 data, “Combo Tv (Nova100)” dengan total Churn & Not Churn 3244 data, “Combo Tv (Sonic150)” dengan total Churn & Not Churn 3238 data, “BASIC” dengan total Churn & Not Churn 3235 data, “Combo Tv (Gamer150)” dengan total Churn & Not Churn 3226 data, “Combo Tv Pakubuwono (Android)” dengan total Churn & Not Churn 3216 data, “Basic Tv” dengan total Churn & Not Churn 3213 data, dan “Combo Tv (Hype75)” dengan total Churn & Not Churn 3199 data.

Gambar 4.33 menampilkan data *Tv Plan* yang mengalami churn. Pada data tersebut ditampilkan jumlah data yang mengalami churn dan persentase dari data churn untuk *Tv Plan* keseluruhan. Sementara itu, Gambar 4.34 menampilkan data *Tv Plan* yang tidak mengalami churn. Pada data tersebut ditampilkan jumlah data yang tidak mengalami churn dan persentase dari data *not churn* untuk *Tv Plan* keseluruhan. Data yang ditampilkan dibatasi hanya 10 baris, agar tampilan pada *website* lebih tertata dan tidak membuang-buang ruang pada *website*.

5. *Advance Promo*

Dari Gambar 4.35 dapat ditarik kesimpulan data sebagai berikut.

Tabel 4.5 Data Rank *Advance Promo*

Data <i>Advance Promo</i>		
Churn / Not Churn	Nama <i>Plan</i>	Ran k
Churn	Promo Disc 20% (12 Bulan) Jet20 Mbps Combo New	1
	Promo Disc 20% (12 Bulan) Jet20 Combo A	2
	Promo Disc 20% (12 Bulan) Value30 Mbps Combo New	3
	Promo Akhir Tahun - 2021 (All Res)	4
	Promo Hari Pahlawan - 2021 (All Res)	5
	Credit Card Payment Reward Discount	6



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Promo Ramadhan Up to 30% 2021 (Fast550 Combo, Gamber150 Mbps, Gamer150 Combo)	7
Promo Ramadhan Up to 30% 2021 (Nova100Mbbps, Nova100 Combo, Sonic150 Combo)	8
Promo Disc 20% (12 Bulan) Fast50 Mbps Combo New	9
Promo Switching 2022 Disc 30% (12 Bulan) Jet20 MMbps Combo New	10
Advance Payment Promo 9+3 (Existing)	1
Credit Card Payment Reward Discount	2
Promo Special Thamrin Residence Disc 20% (6 Bulan) Fast50 Mbps New	3
Discount 10000 - PI21 (12Months)	4
Not Churn Disc Demoline PoP 50 Mbps	5
3 Month Advance Pament	6
3 Month Advance Pament for selected vendor	7
Advance Payment Promo 9+3 (New SA)	8
Advance Payment 6 Mmonth	9
Discount 15000 - PI21 (11months)	10

Terlihat untuk *Advance Promo* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Advance Promo* “Promo Disc 20% (12 Bulan) Jet20 Mbps Combo New”. Sementara itu, *Advance Promo* yang tidak mengalami churn paling banyak yaitu berada pada *Advance Promo* “Advance Payment Promo 9+3 (Existing)”.

Gambar 4.37 menampilkan 10 data *Advance Promo*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*. Pada Gambar 4.35 di dapat terlihat bahwa data sulit untuk dibaca. Oleh karena itu, diperlukannya penampilan data dengan sistem *ranking*. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Di sini akan



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

ditampilkan data “Top 5” dan “Top 10” data *Advance Promo*. Data tersebut ditampilkan pada Gambar 4.38 dan Gambar 4.40.

Berdasarkan Gambar 4.38 dan Gambar 4.39, dapat ditarik kesimpulan bahwa data yang memiliki jumlah data gabungan antara Churn dan Not Churn 5 teratas antara lain, "Credit Card Payment Reward Discount", "Advance Payment Promo 9+3 (Existing)", "Promo Special Thamrin Residence Disc 20% (6 Bulan) Fast50 Mbps New", "Disc 10000 - PI21 (12Months)", dan "Disc Demoline PoP 50 Mbps". "Credit Card Payment Reward Discount" dengan total Churn & Not Churn 3340 data, "Advance Payment Promo 9+3 (Existing)" dengan total Churn & Not Churn 3324 data, "Promo Special Thamrin Residence Disc 20% (6 Bulan) Fast50 Mbps New" dengan total Churn & Not Churn 3240 data, "Disc 10000 - PI21 (12Months)" dengan total Churn & Not Churn 3219 data, dan "Disc Demoline PoP 50 Mbps" dengan total Churn & Not Churn 3211 data. Berdasarkan Gambar 4.40 dan Gambar 4.41, dapat ditarik kesimpulan bahwa data yang memiliki jumlah data gabungan antara Churn dan Not Churn 10 teratas antara lain, "Credit Card Payment Reward Discount", "Advance Payment Promo 9+3 (Existing)", "Promo Special Thamrin Residence Disc 20% (6 Bulan) Fast50 Mbps New", "Disc 10000 - PI21 (12Months)", "Disc Demoline PoP 50 Mbps", "3 Month Advance Payment", "3 Month Advance Payment for selected vendor", "Advance Payment Promo 9+3 (New SA)", "Discount 15000 - PI21 (11months)", dan "Advance Payment 6 Month". "Credit Card Payment Reward Discount" dengan total Churn & Not Churn 3340 data, "Advance Payment Promo 9+3 (Existing)" dengan total Churn & Not Churn 3324 data, "Promo Special Thamrin Residence Disc 20% (6 Bulan) Fast50 Mbps New" dengan total Churn & Not Churn 3240 data, "Disc 10000 - PI21 (12Months)" dengan total Churn & Not Churn 3219 data, "Disc Demoline PoP 50 Mbps" dengan total Churn & Not Churn 3211 data, "3 Month Advance Payment" dengan total Churn & Not Churn 3176 data, "3 Month Advance Payment for selected vendor" dengan total Churn & Not Churn 3172 data, "Advance Payment Promo 9+3 (New SA)" dengan total Churn & Not Churn 3171 data, "Discount 15000 - PI21 (11months)" dengan total Churn & Not Churn 3163 data, dan "Advance Payment 6 Month" dengan total Churn & Not Churn 3162 data.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Gambar 4.42 menampilkan data *Advance Promo* yang mengalami churn. Pada data tersebut ditampilkan jumlah data yang mengalami churn dan persentase dari data churn untuk *Advance Promo* keseluruhan. Sementara itu Gambar 4.43 di atas menampilkan data *Advance Promo* yang tidak mengalami churn. Pada data tersebut ditampilkan jumlah data yang tidak mengalami churn dan persentase dari data *not churn* untuk *Advance Promo* keseluruhan. Data yang ditampilkan dibatasi hanya 10 baris, agar tampilan pada *website* lebih tertata dan tidak membuang-buang ruang pada *website*.

B. *Dataset 2*

Dataset 2 terdiri atas beberapa jenis pengelompokan, antara lain:

1. Keseluruhan Data

Pada *website* jika dilihat di grafik pie terdapat sedikit perbedaan pada bagian kalkulasi dari persentase yang dihasilkan dari data Not Churn. Hal ini dikarenakan, pada grafik pie terjadi pembulatan yang mengakibatkan angka dari nilai persentase ini berubah menjadi 3.0%.

Tabel 4.6 Rasio Data Churn & Not Churn

Rasio Data		
Churn / Not Churn	Jumlah Data	Persentase
Churn	1455	2,96%
Non	47712	97,04%

Berdasarkan tabel di atas didapatkan hasil jumlah data yang mengalami churn sebanyak 1.455 data dengan nilai persentase sebesar 2,96%. Sementara itu, data yang tidak mengalami churn sebanyak 47.712 data dengan nilai persentase sebesar 97,04%.

Meninjau dari *dataset* yang dibuat, *dataset* yang melalui prediksi memuat data churn sebanyak 1.475. *Dataset 2* ini tidak mengalami perubahan jumlah data churn yang dimasukkan dari *dataset 1* dari pengujian sebelumnya. Oleh karena itu, dapat diketahui bahwa, jumlah data churn hasil prediksi churn yang dilakukan web



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

bernilai sama dengan jumlah data churn hasil prediksi sebelumnya. Dengan rasio sebesar $\pm 8,1:10$ atau 0,8142372881 dan tingkat *error* berada di angka 0,1857627119.

Berbeda halnya jika ditinjau hasil dari uji *dataset* 1 dan uji *dataset* 2 ini. Pada *dataset* 1, data *Area Name*, *Plan*, *Tv Plan*, dan *Advance Promo* dibatasi sebanyak 15 *unique value*. Sementara itu pada *dataset* 2 ini, data *Area Name*, *Plan*, *Tv Plan*, dan *Advance Promo* tidak dibatasi *unique value* dari masing-masing data. Hal ini berdampak pada hasil prediksi. Pada Hasil pengujian menggunakan *dataset* 1 didapatkan jumlah data yang mengalami churn sebanyak 1.201 data, sementara itu hasil pengujian menggunakan *dataset* 2 didapatkan jumlah data yang mengalami churn sebanyak 1.455 data. Dengan ini, dapat diketahui bahwa hasil dari data *Area Name*, *Plan*, *Tv Plan*, dan *Advance Promo* yang tidak dibatasi ini, menghasilkan data churn sebanyak 254 data.

2. *Area Name*

Berdasarkan pengujian yang dilakukan, dapat ditarik kesimpulan data sebagai berikut .

Tabel 4.7 Data Rank *Area Name*

Data <i>Area Name</i>		
Churn / Not Churn	Nama Area	Rank
Churn	Jakarta	1
	Tangerang	2
	Bandung	3
	Bekasi	4
	Cibubur	5
	Palembang	6
	Depok	7
	Medan	8
	Cirebon	9
	Surabaya	10
Not Churn	Jakarta	1



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Medan	2
Lampung	3
Purwokerto	4
Bekasi	5
Depok	6
Semarang	7
Tangerang	8
Malang	9
Surabaya	10

Terlihat untuk *Area Name* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Area Name* Jakarta. Sementara itu, *Area Name* yang tidak mengalami churn paling banyak yaitu berada pada *Area Name* Jakarta.

Grafik batang proporsi data churn & tidak churn pada *Area Name* terlihat bahwa “Jakarta” pada variabel churn sangatlah memndominasi dibandingkan dengan *Area Name* lainnya. Tabel proporsi data churn & tidak churn pada *Area Name* menampilkan 10 ddata *Area Name*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*. Grafik batang ini terlihat cukup sulit untuk dibaca, oleh karena itu dibuatkanlah sistem *ranking* untuk penampilan data hasilnya. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Data ini akan dibagi menjadi Top 5 dan Top 10.

Pada grafik batang Top 5 dan Top 10 yang dibuat dapat ditarik kesimpulan yang akan ditampilkan pada Tabel 4.8 berikut.

Tabel 4.8 Data Kesimpulan Top 5 & Top 10 *Area Name*

Feature	Variabel	Total Data	Rank
<i>Area Name</i>	Jakarta	2612	1
	Tangerang	2160	2
	Medan	2142	3



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Bekasi	2118	4
Depok	2092	5
Lampung	2067	6
Bandung	2061	7
Semarang	2060	8
Palembang	2058	9
Surabaya	2053	10

Selanjutnya pada tabel persentase data yang mengalami churn dan tidak mengalami churn pada *Area Name* dapat dilihat pada L-11.

3. *Plan*

Berdasarkan pengujian yang dilakukan, dapat ditarik kesimpulan data sebagai berikut .

Tabel 4.9 Data Rank *Plan*

Data Plan		
Churn / Not Churn	Nama Plan	Rank
Churn	Jet20 Combo New	1
	Jet20 Combo Internet A	2
	Value30 Combo New	3
	Value30 Combo	4
	Fast50 Combo New	5
	Fast50 Combo	6
	Nova100 Combo	7
	Business Pro 100 Mbps	8
	Hype75 Combo Nnew	9
	Value30 Combo Internet A	10
Not Churn	Jet20 Combo New	1
	Value30 Combo New	2
	Nova 100 Mbps+ XTRA	3
	Jet20 Mbps	4


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Supernova300 ComboTv Pakubuwono	5
Jet 20 Mbps+	6
Gamer150 Mbps Reg (PA)	7
Jet20 Combo Internet A	8
Gamer150 (PI-22C)	9
Nova 100 Internet (AP77)	10

Terlihat untuk *Plan* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Plan* “Jet20 Combo New”. Sementara itu, *Plan* yang tidak mengalami churn paling banyak yaitu berada pada *Plan* “Jet20 Combo New”.

Tabel proporsi data churn & tidak churn pada *Plan* menampilkan 10 data *Plan*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang-buang ruang pada *website*. Grafik batang ini terlihat cukup sulit untuk dibaca, oleh karena itu dibuatkanlah sistem *ranking* untuk penampilan data hasilnya. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Data ini akan dibagi menjadi Top 5 dan Top 10.

Pada grafik batang Top 5 dan Top 10 yang dibuat dapat ditarik kesimpulan yang akan ditampilkan pada Tabel 4.10 berikut.

 Tabel 4.10 Data Kesimpulan Top 5 & Top 10 *Plan*

Feature	Variabel	Total Data	Rank
	Jet20 Combo New	793	1
	Jet20 Combo Internet A	275	2
<i>Plan</i>	Value30 Combo New	260	3
	Value30 Combo	191	4
	Fast50 Combo New	179	5
	Nova100 Combo	144	6



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Nova 100 Mbps+ XTRA	142	7
Jet 20 Mbps+	136	8
Jet20 Mbps	133	9
Gamer150 (PI-22C)	133	10

Selanjutnya pada tabel persentase data yang mengalami churn dan tidak mengalami churn pada *Plan* dapat dilihat pada L-12.

4. *Tv Plan*

Berdasarkan pengujian yang dilakukan, dapat ditarik kesimpulan data sebagai berikut .

Tabel 4.11 Data Rank *Tv Plan*

Data <i>Tv Plan</i>		
Churn / Not Churn	Nama <i>Plan</i>	Rank
Churn	Combo Tv (jet20)	1
	Jet20 Mbps (ComboTv)	2
	Combo Tv (Value30)	3
	Combo Tv (Fast50)	4
	SmarTv+	5
	Combo Tv (Nova100)	6
	Value30 Mbps (ComboTv + ELK)	7
	Combo Tv (Hyppe75)	8
	Advance Tv	9
	StarTv	10
Not Churn	Combo Tv (Jet20)	1
	ComboTv Pakubuwono	2
	Jet20 Mbps (ComboTv)	3
	Combo Tv (Value30)	4
	Value30 Mbps (ComboTv + ELK)	5


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

SOHO A	6
STAR	7
Basic Tv	8
ComboTv Pakubuwono (Android)	9
Xtra ComboTv (Jet20)	10

Terlihat untuk *Tv Plan* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Tv Plan* “Combo Tv (jet20)”. Sementara itu, *Tv Plan* yang tidak mengalami churn paling banyak yaitu berada pada *Tv Plan* “Combo Tv (Jet20)”.

Tabel proporsi data churn & tidak churn pada *Tv Plan* menampilkan 10 data *Tv Plan*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*. Grafik batang ini terlihat cukup sulit untuk dibaca, oleh karena itu dibuatkanlah sistem *ranking* untuk penampilan data hasilnya. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Data ini akan dibagi menjadi Top 5 dan Top 10.

Pada grafik batang Top 5 dan Top 10 yang dibuat dapat ditarik kesimpulan yang akan ditampilkan pada Tabel 4.12 berikut.

 Tabel 4.10 Data Kesimpulan Top 5 & Top 10 *Tv Plan*

Feature	Variabel	Total Data	Rank
	Combo <i>Tv</i> (Jet20)	1495	1
	Jet20 Mbps (Combo <i>Tv</i>)	1252	2
<i>Tv Plan</i>	Combo <i>Tv</i> (Value30)	1183	3
	Combo <i>Tv</i> (Fast50)	1019	4
	Value30 Mbps (Combo <i>Tv</i> +ELK)	1003	5
	Combo <i>Tv</i> Pakubuwono	988	6


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

STAR	981	7
SOHO A	980	8
Basic Tv	978	9
SmarTv	970	10

Selanjutnya pada tabel persentase data yang mengalami churn dan tidak mengalami churn pada *Tv Plan* dapat dilihat pada L-13.

 5. *Advance Promo*

Berdasarkan pengujian yang dilakukan, dapat ditarik kesimpulan data sebagai berikut.

 Tabel 4.13 Data Rank *Advance Promo*

Data <i>Advance Promo</i>		
Churn / Not Churn	Nama <i>Plan</i>	Rank
Churn	Promo Disc 20% (12 Bulan) Jet20 Mbps Combo New	1
	Promo Disc 20% (12 Bulan) Jet20 Combo A	2
	Promo Disc 20% (12 Bulan) Value30 Mbps Combo New	3
	Promo Akhir Tahun - 2021 (All Res)	4
	Promo Hari Pahlawan - 2021 (All Res)	5
	Credit Card Payment Reward Discount	6
	Promo Ramadhan Up to 30% 2021 (Fast50 Combo, Gamber150 Mbps, Gamer150 Combo)	7
	Promo Ramadhan Up to 30% 2021 (Nova100 Mbps, Nova100 Combo, Sonic150 Combo)	8
	Promo Disc 20% (12 Bulan) Fast550 Mbps Combo New	9
	RWB Promo Diskon 50% X 6	10

	Promo Disc 20% (12 Bulan) Jet20 Mbps Combo New	1
	Promo Disc 20% (12 Bulan) Value30 Mbps Combo New	2
	Promo CLBK Disc 25% (12 Bulan) Jet20 Internet	3
	Promo Akhir Tahun - 2021 (Gamer Combo)	4
	Promo WOW FLASH SALE Fast50 Combo A	5
	Promo CLBK 25% (12 Bulan) Business20 Mbps	6
Not Churn	Promo Disc 20% (12 Bulan) Jet20 Combo A	7
	Promo New RFS HRB Disc 20% (6 Bulan) Nova100 Combo A	8
	Promo Shocktober Disc 30% (12 Bulan) MyGamer250 Combo A	9
	Promo Ultah MyRep Disc 20% + 7% (12 Bulan) Fast50 Mbps New	10

Terlihat untuk *Advance Promo* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Advance Promo* “Promo Disc 20% (12 Bulan) Jet20 Mbps Combo New”. Sementara itu, *Advance Promo* yang tidak mengalami churn paling banyak yaitu berada pada *Advance Promo* “Promo Disc 20% (12 Bulan) Jet20 Mbps Combo New”.

Tabel proporsi data churn & tidak churn pada *Tv Plan* menampilkan 10 data *Advance Promo*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*. Grafik batang ini terlihat cukup sulit untuk dibaca, oleh karena itu dibuatkanlah sistem *ranking* untuk penampilan data hasilnya. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Data ini akan dibagi menjadi Top 5 dan Top 10.

Pada grafik batang Top 5 dan Top 10 yang dibuat dapat ditarik kesimpulan yang akan ditampilkan pada Tabel 4.12 berikut.

Tabel 4.14 Data Kesimpulan Top 5 & Top 10 *Advance Promo*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, pennisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengunumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Feature	Variabel	Total Data	Rank
<i>Advance Promo</i>	Promo Disc 20% (12 Bulan) Jet20 Mbps Combo New	763	1
	Promo Disc 20% (12 Bulan) Jet20 Combo A	294	2
	Promo Disc 20% (12 Bulan) Value30 Mbps Combo New	262	3
	Promo Akhir Tahun - 2021 (All Res)	202	4
	Credit Card Payment Reward Discount	174	5
	Promo Disc 20% (12 Bulan) Fast50 Mbps Combo New	168	6
	Promo Hari Pahlawan - 2021 (All Res)	164	7
	Prommo CLBK Disc 25% (12 Bulan) Jet20 Internet	159	8
	Promo Akhir Tahun - 2021 (Gamber Combo)	159	9
	Promo Ramadhan Up to 30% 2021 (Fast Combo, Gamer 150 Mbps, Gamer150 Combo)	158	10

Selanjutnya pada tabel persentase data yang mengalami churn dan tidak mengalami churn pada *Tv Plan* dapat dilihat pada L-14.

C. *Dataset 3*

Dataset 3 terdiri atas beberapa jenis pengelompokkan, antara lain:

1. Keseluruhan Data

Pada *website* jika dilihat di grafik pie terlihat bahwa untuk bagian Churn mendapatkan nilai 0,5%. Sementara itu pada bagian Not Churn bernilai 99,5%.

Tabel 4.15 Rasio Data Churn & Not Churn

Rasio Data		
Churn / Not Churn	Jumlah Data	Persentase



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Churn	703	0,50%
Not Churn	139297	99,50%

Berdasarkan tabel di atas, model *machine learning* dapat memprediksi jumlah data yang mengalami churn sebanyak 703 data dengan nilai persentase sebesar 0,50%. Sementara itu, data yang tidak mengalami churn sebanyak 139.297 data dengan nilai persentase sebesar 99,50%.

2. *Area Name*

Berdasarkan pengujian yang dilakukan, dapat ditarik kesimpulan data sebagai berikut.

Tabel 4.16 Rank *Area Name*

Data Area Name		
Churn / Not Churn	Nama Area	Rank
Churn	Cilegon	1
	Cirebon	2
	Bogor	3
	Bekassi	4
	Cibubur	5
	Bandung	6
	Karawang	7
	Lampung	8
	Jambi	9
	Makassar	10
Not Churn	Bali	1
	Serang	2
	Cibubur	3
	Tegal	4
	Palembang	5
	Surabaya	6



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Depok	7
Jakarta	8
Jambi	9
Pekanbaru	10

Terlihat untuk *Area Name* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Area Name* “Cilegon”. Sementara itu, *Area Name* yang tidak mengalami churn paling banyak yaitu berada pada *Area Name* “Bali

Tabel proporsi data churn & tidak churn pada *Area Name* menampilkan 10 data *Area Name*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang-buang ruang pada *website*. Grafik batang ini terlihat cukup sulit untuk dibaca, oleh karena itu dibuatkanlah sistem *ranking* untuk penampilan data hasilnya. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Data ini akan dibagi menjadi Top 5 dan Top 10.

Pada grafik batang Top 5 dan Top 10 yang dibuat dapat ditarik kesimpulan yang akan ditampilkan pada Tabel 4.17 berikut.

Tabel 4.17 Data Kesimpulan Top 5 & Top 10 *Area Name*

Feature	Variabel	Total Data	Rank
	Bali	5937	1
	Cibubur	5925	2
	Serang	5915	3
	Palembang	5883	4
<i>Area Name</i>	Tegal	5882	5
	Surabaya	5880	6
	Jambi	5876	7
	Depok	5872	8
	Bekasi	5863	9



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Jakarta 5860 10

3. *Plan*

Berdasarkan pengujian yang dilakukan, dapat ditarik kesimpulan data sebagai berikut.

Tabel 4.18 Rank *Plan*

Data <i>Plan</i>		
Churn / Not Churn	Nama <i>Plan</i>	Rank
Churn	Business 75 (PI20)	1
	Business Pro 500 (PI20)	2
	Business Pro 100 Mbps	3
	Business Pro 150 (PI20)	4
	Business Pro 150 (PI21)	5
	Business Pro 150 Mbps	6
	Business 50 (PI20)	7
	Business 50 Mbps	8
	Business 50 (PI21)	9
	Business 300 Mbps (free <i>installation fee</i>)	10
Not Churn	Nova.100+	1
	Basic 30 Mbps	2
	SNAP (Internet Up To 10 Mbps)	3
	Supernova300 Mbps (PA)	4
	Fast50 Mbps	5
	Magic Wifi 200 (PI21)	6
	IBS.CA Business Pro 150 Mbps	7
	Nova100 (PI-22E)	8
	VIP NRO 2021+	9
	Value30 (PI-22B)	10


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Terlihat untuk *Plan* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Plan* “ Business 75 (PI20)”. Sementara itu, *Plan* yang tidak mengalami churn paling banyak yaitu berada pada *Plan* “Nova.100+”.

Tabel proporsi data churn & tidak churn pada *Plan* menampilkan 10 data *Plan*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*. Grafik batang ini terlihat cukup sulit untuk dibaca, oleh karena itu dibuatkanlah sistem *ranking* untuk penampilan data hasilnya. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Data ini akan dibagi menjadi Top 5 dan Top 10.

Pada grafik batang Top 5 dan Top 10 yang dibuat dapat ditarik kesimpulan yang akan ditampilkan pada Tabel 4.19 berikut.

 Tabel 4.19 Data Kesimpulan Top 5 & Top 10 *Plan*

Feature	Variabel	Total Data	Rank
<i>Plan</i>	Nova.100+	369	1
	Basic 30 Mbps	368	2
	SNAP (Internet Up To 10 Mbps)	363	3
	Supernova300 Mbps (PA)	362	4
	Fast50 Mbps	362	5
	Magic Wifi 200 (PI21)	356	6
	IBS.CA Business Pro 150 Mbps	353	7



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Nova100 (PI-22E)	353	8
VIP NRO 2021+	352	9
Value30 (PI-22B)	352	10

4. *Tv Plan*

Berdasarkan pengujian yang dilakukan, dapat ditarik kesimpulan data sebagai berikut.

Tabel 4.20 Rank *Tv Plan*

Data <i>Tv Plan</i>		
Churn / Not Churn	Nama <i>Tv Plan</i>	Rank
Churn	Advance <i>Tv</i> (Android)	1
	Advance <i>Tv</i>	2
	BASIC	3
	Smar <i>Tv</i> +	4
	Star <i>Tv</i> Plus (Android)	5
	Star <i>Tv</i> Plus	6
	SOHO A	7
	<i>Tv</i> SOHO B (Movies dan Sports)	8
	Smar <i>Tv</i>	9
	Star <i>Tv</i> SOHO	10
Not Churn	Star <i>Tv</i> Plus	1
	Tanpa <i>Plan</i>	2
	Local Channel (Android)	3
	Combo <i>Tv</i> (Sonic150)	4
	Xtra Combo <i>Tv</i> (V30 + ELK)	5
	Star <i>Tv</i> Plus (Android)	6
	SOHO A	7
	Smar <i>Tv</i> +	8



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

ComboTv Pakubuwono (Android)	9
STAR	10

Terlihat untuk *Tv Plan* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Tv Plan* “Advance *Tv* (Android)”. Sementara itu, *Tv Plan* yang tidak mengalami churn paling banyak yaitu berada pada *Tv Plan* “Star *Tv* Plus”.

Tabel proporsi data churn & tidak churn pada *Tv Plan* menampilkan 10 data *Tv Plan*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang–buang ruang pada *website*. Grafik batang ini terlihat cukup sulit untuk dibaca, oleh karena itu dibuatkanlah sistem *ranking* untuk penampilan data hasilnya. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Data ini akan dibagi menjadi Top 5 dan Top 10.

Pada grafik batang Top 5 dan Top 10 yang dibuat dapat ditarik kesimpulan yang akan ditampilkan pada Tabel 4.21 berikut.

Tabel 4.21 Data Kesimpulan Top 5 & Top 10 *Tv Plan*

Feature	Variabel	Total Data	Rank
	Star <i>Tv</i> Plus	2871	1
	Local Channel (Android)	2841	2
	Tanpa <i>Plan</i>	2838	3
	Combo <i>Tv</i> (Sonic150)	2835	4
<i>Tv Plan</i>	Star <i>Tv</i> Plus (Android)	2809	5
	Smar <i>Tv</i> +	2802	6
	SOHO A	2796	7
	Xtra Combo <i>Tv</i> (V30 + ELK)	2789	8



- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
 2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

STAR	2788	9
StarTv	2779	10

5. Advance Promo

Berdasarkan pengujian yang dilakukan, dapat ditarik kesimpulan data sebagai berikut.

Tabel 4.22 Rank Advance Promo

Data Advance Promo		
Churn / Not Churn	Nama Advance Promo	Rank
Churn	RWB Promo Diskon 50% X 6	1
	Tanpa Advance Promo	2
	Upsell Discount (20% 3 Months)	3
	RWB Promo Diskon 35% X 6	4
	RWB Promo Diskon 35%	5
	Promo Disc 25% (12 Bulan) Business300 Mbps	6
	RWB Promo Diskon 50% X 3	7
	Promo CLBK Disc 25% (12 Bulan) Hype73 Mbps New	8
	Promo Harbolnas September Nova100 Combo A	9
	Promo Disc 20% (12 Bulan) BusinessPro500 Mbps	10
Not Churn	Promo Ruko Special Price Disc 15% (6 Bulan) Jet20 Combo A	1
	Promo Nasional 2023 (PA 5 Get 1) Fast50 Combo Internet A	2
	Promo Oktober 2022 Disc 25% (12 Bulan) Nova100 Combo A	3



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Promo Nasional 2023 (PA 5 Get 1) Value30 Combo Internet A	4
Promo Merdeka Disc 25% 6 Bulan (Nova100)	5
Promo Upsell Disc 15% (6 Bulan) Nova100 Combo A	6
Promo Disc 20% (12 Bulan) BusinessPro150 Mbps	7
Promo Switching 2022 Disc 30% (12 Bulan) Fast50 Mbpps New	8
Advance Payment Promo 9+3 (New SA)	9
Promo New RFS Disc 15% Jet20 Mbpps Combo New	10

Terlihat untuk *Advance Promo* yang memberikan sumbangsi terbesar untuk data yang mengalami churn, yaitu *Advance Promo* “Advance Tv (Android)”. Sementara itu, *Advance Promo* yang tidak mengalami churn paling banyak yaitu berada pada *Advance Promo* “Star Tv Plus”.

Tabel proporsi data churn & tidak churn pada *Advance Promo* menampilkan 10 data *Advance Promo*. Pembatasan pada penampilan data ini dilakukan supaya tampilan pada *website* lebih tertata dan tidak membuang-buang ruang pada *website*. Grafik batang ini terlihat cukup sulit untuk dibaca, oleh karena itu dibuatkanlah sistem *ranking* untuk penampilan data hasilnya. *Ranking* ini akan memuat data dengan jumlah “Churn” dan “Not Churn” yang digabungkan. Selanjutnya data tersebut akan diurutkan dari jumlah data paling banyak hingga ke data dengan jumlah paling sedikit. Data ini akan dibagi menjadi Top 5 dan Top 10.

Pada grafik batang Top 5 dan Top 10 yang dibuat dapat ditarik kesimpulan yang akan ditampilkan pada Tabel 4.23 berikut.

Tabel 4.23 Data Kesimpulan Top 5 & Top 10 *Advance Promo*

Feature	Variabel	Total Data	Rank
---------	----------	------------	------



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Promo Ruko Special Price Disc 15% (6 Bulan) Jet20 Combo A	430	1
Promo Nasional 2023 (PA 5 Get 1) Fast50 Combo Internet A	420	2
Promo Oktober 2022 Disc 25% (12 Bulan) Nova100 Combo A	419	3
Promo Merdeka Disc 25% 6 Bulan (Nova100)	418	4
Promo Nasional 2023 (PA 5 Get 1) Value30 Combo Internet A	418	5
Promo Upsell Disc 15% (6 Bulan) Nova100 Combo A	415	6
Promo Disc 20% (12 Bulan) BusinessPro150 Mbps	413	7
Advance Payment Promo 9+3 (New SA)	409	8
Promo Switching 2022 Disc 30% (12 Bulan) Fast50 Mbps New	408	9
Promo New RFS Disc 15% Jet20 Mbps Combo New	408	10

Advance Promo

4.1.3 Pengujian Dengan Data *Real*

4.1.3.1 Deskripsi Pengujian

Pengujian model yang telah dibuat dengan menggunakan data *real* ini bertujuan untuk melihat seberapa baik model dapat mengidentifikasi data yang mengalami churn. Pengujian ini hanya akan menggunakan data – data pelanggan yang sudah mengalami churn.

4.1.3.2 Prosedur Pengujian

Berikut adalah prosedur untuk pengujian model dengan menggunakan data *real* sebagai *dataset* ujinya.



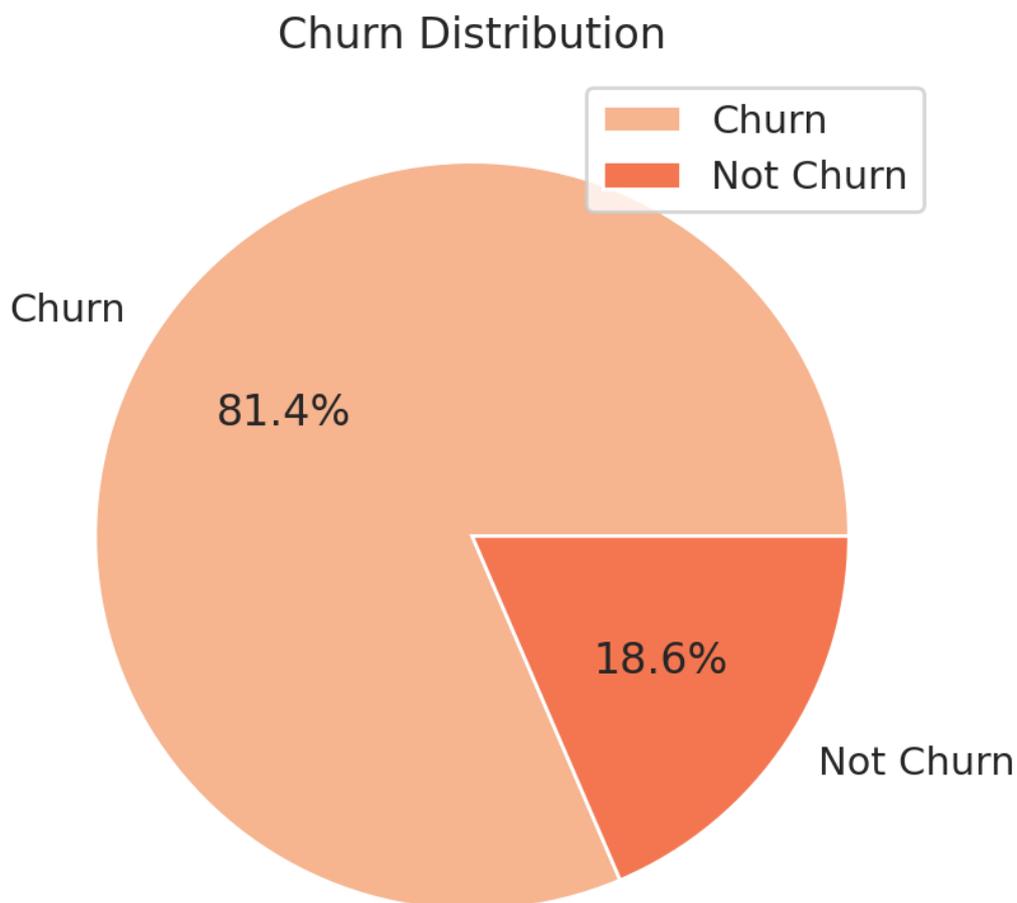
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

1. Menyiapkan *dataset*. *Dataset* yang disiapkan ini mengambil dari data perusahaan. Data. *Dataset* ini memuat data pelanggan yang sudah dikatakan churn atau berhenti berlangganan. *Dataset* ini memuat jumlah data churn sebanyak 1.475 data.
2. *Upload dataset* pada *website* yang telah dibuat
3. Tinjaulah hasil yang diberikan oleh *website* prediksi churn yang telah dibuat

4.1.3.3 Data Hasil Pengujian

Pengujian ini dilakukan dengan cara memprediksi *dataset* dengan menggunakan sistem churn prediksi yang telah dibuat. Setelah proses prediksi selesai dilakukan oleh *website*, selanjutnya meninjau hasil. Pengujian ini hanya berfokus pada hasil jumlah churn dan not churn data setelah diprediksi.



Gambar 4.45 Grafik Pie Distribusi Churn Seluruh Data



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

	Jenis Churn	Jumlah Data	Persentase
0	Churn	1201	81.42%
1	Not Churn	274	18.58%

Gambar 4.46 Hasil Tabel Distribusi Churn Seluruh Data

4.1.3.4 Analisa Hasil Pengujian

Setelah melalui proses prediksi, mendapatkan hasil data Churn sebanyak 1.201 data dari total data sebanyak 1.475 data. Dengan ini data yang mengalami churn hanya sebanyak 81,42%. Hal ini berarti terdapat *error* sebesar 18,58% dari total data yang seharusnya mengalami Churn semua. Berdasarkan hasil tersebut, model *machine learning* yang telah dibuat, dapat dikategorikan baik dalam proses prediksi. Model dikatakan baik dikarenakan ketepatan prediksi masih dapat menyentuh angka 80%.

4.2 Pengujian Generate Churn Report

4.2.1 Deskripsi Pengujian

Pengujian *generate churn report* ini dilakukan untuk melihat seberapa baik sistem yang ada pada *website* prediksi churn dalam membentuk sebuah *file* PDF yang berisi kumpulan grafik yang tampil sesuai dengan yang ada pada *website*. Pengujian ini dilakukan dengan mencoba 3 *dataset dummy* yang telah disiapkan sebelumnya.

4.2.2 Prosedur Pengujian

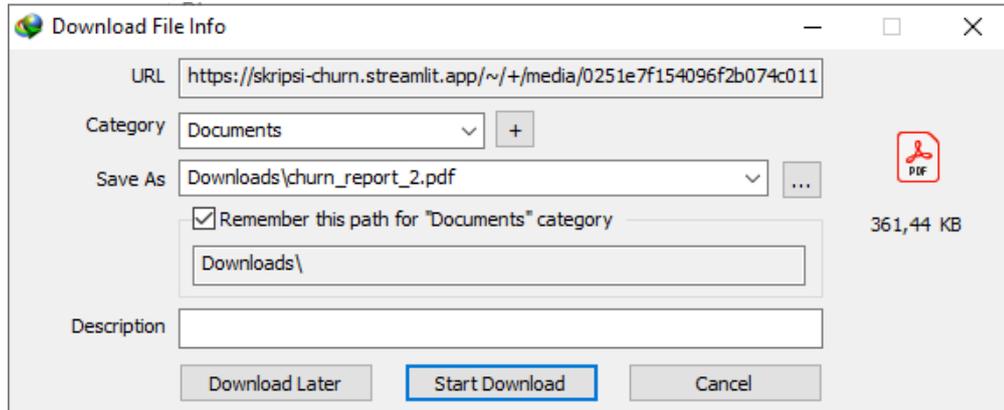
Berikut adalah prosedur pengujian *generate churn report*:

1. Unggah *file dataset*
2. Tunggu *website* memproses semuanya
3. Tekan tombol “Click Here” pada bagian Download Churn Report
4. Tekan tombol “Download Churn Report”, sesaat setelah meng-klik tombol “Click Here”
5. Ulangi langkah 1 hingga 4 dengan menggunakan *dataset dummy* yang berbeda

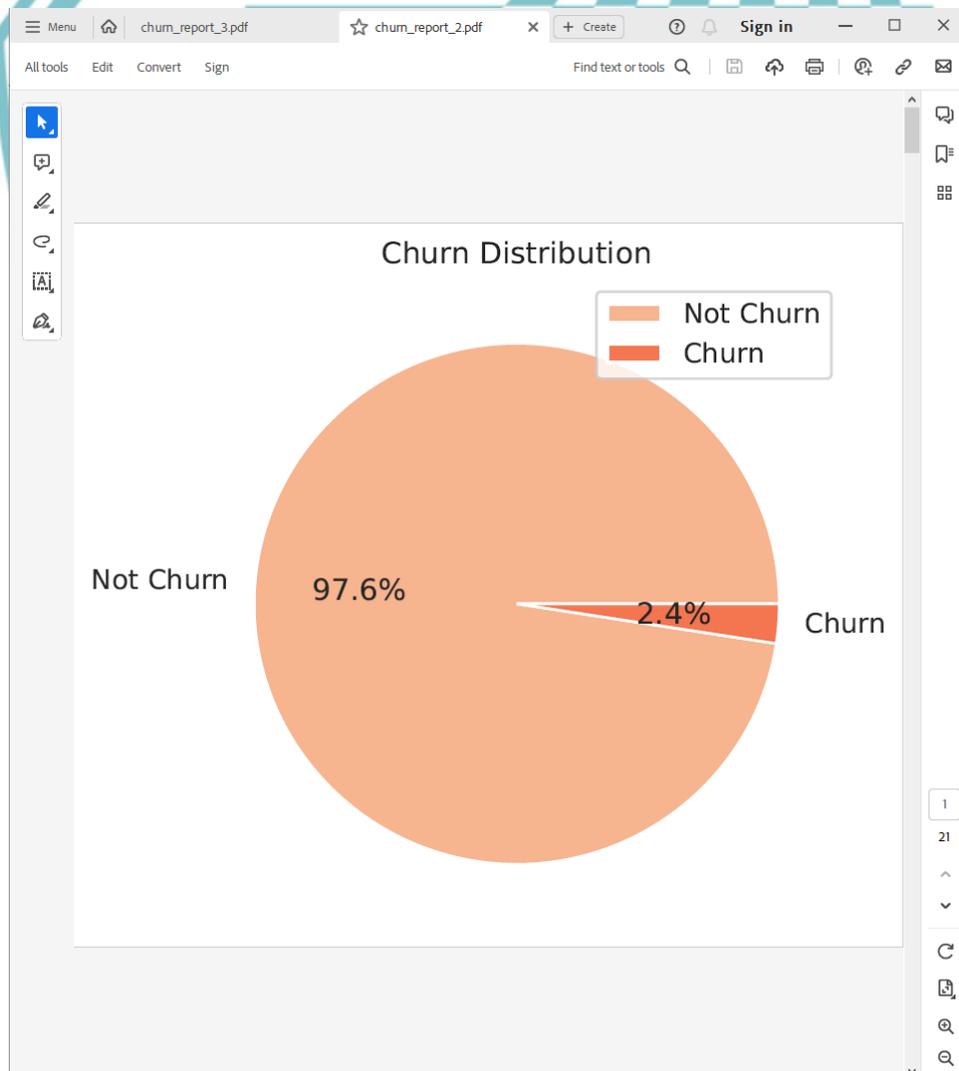


4.2.3 Data Hasil Pengujian

Pengujian *generate churn report* yang telah dilakukan akan menghasilkan seperti pada Gambar 4.47.



Gambar 4.47 File PDF *Dataset Dummy 1* Berhasil *Generated*



Gambar 4.48 Hasil *Churn Report Dataset Dummy 1*

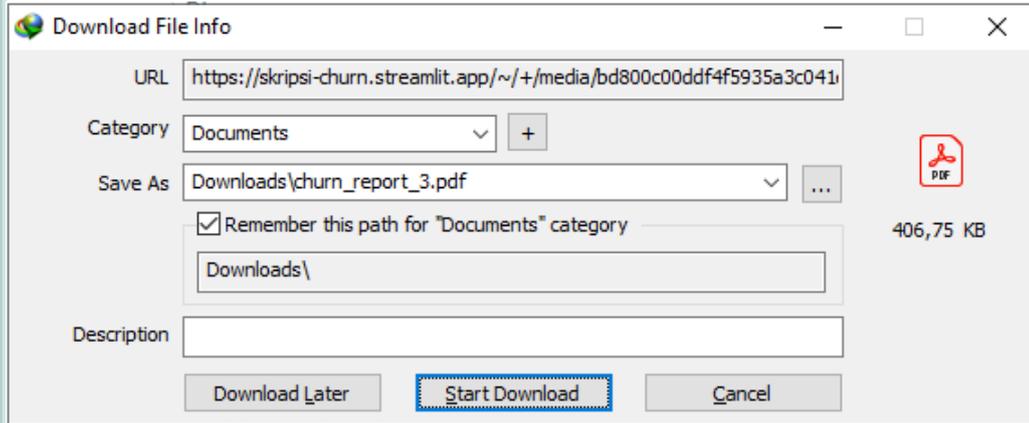
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

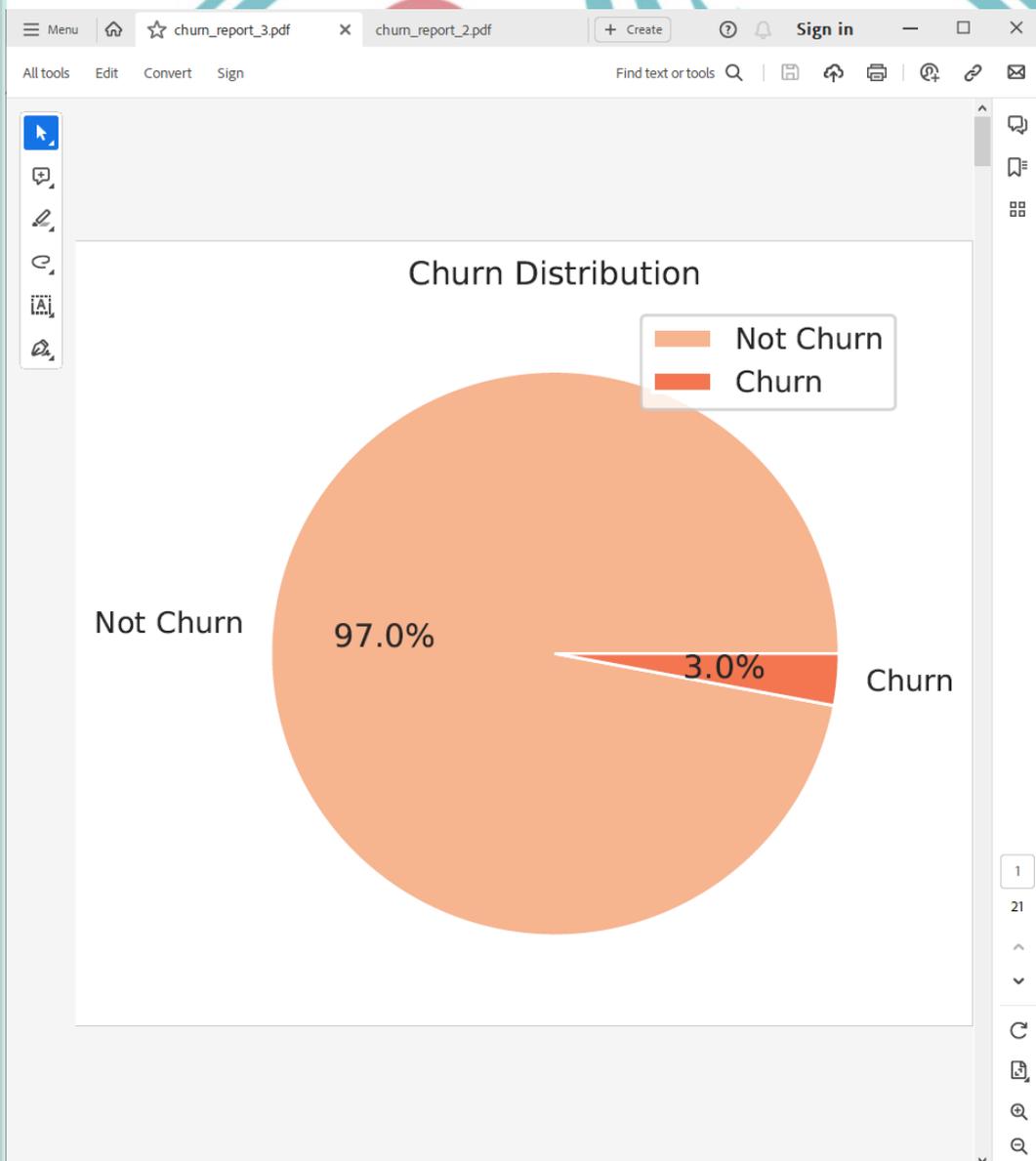


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.49 File PDF *Dataset Dummy 2* Berhasil Generated

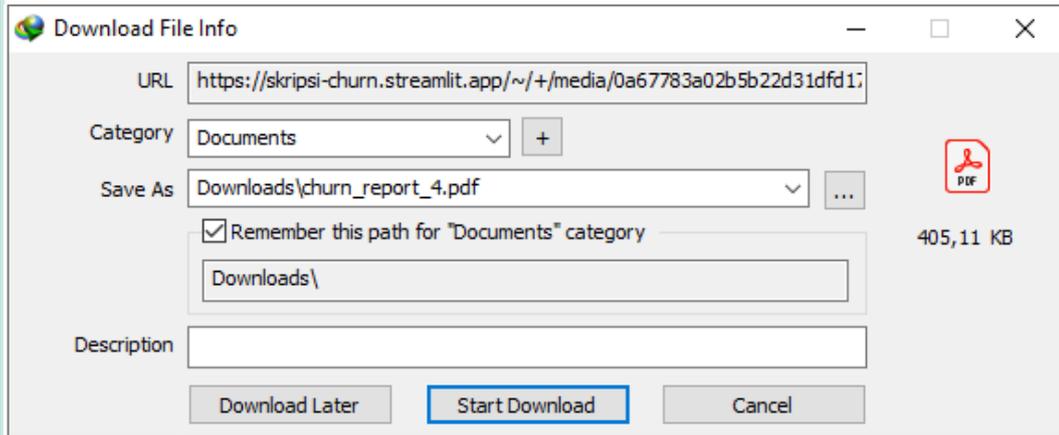


Gambar 4.50 Hasil *Churn Report Dataset Dummy 2*

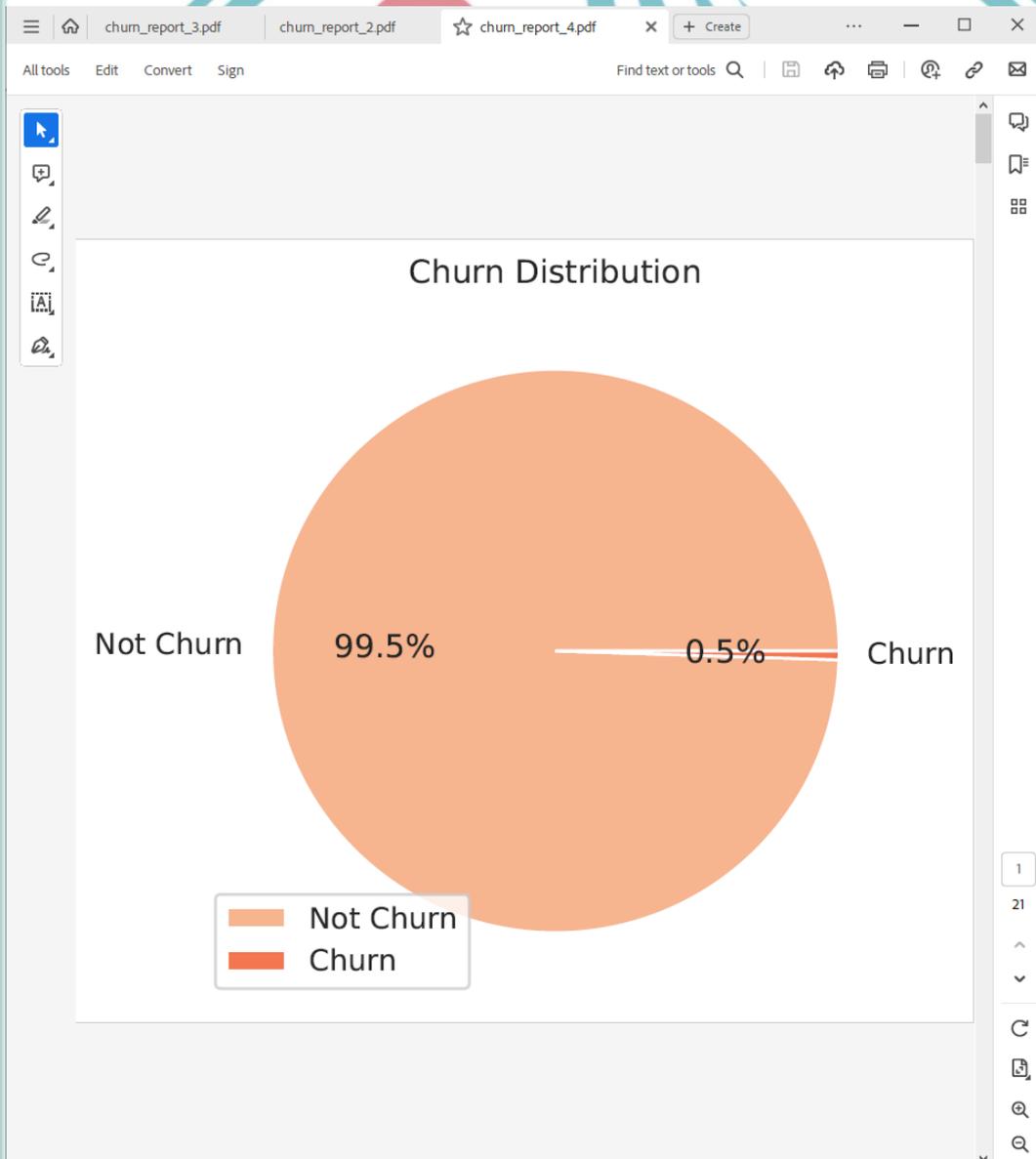


Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 4.51 File PDF *Dataset Dummy 3* Berhasil *Generated*



Gambar 4.52 Hasil *Churn Report Dataset Dummy 3*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

4.2.4 Analisa Hasil Pengujian

Berdasarkan data hasil pengujian *generate churn report* yang telah dilakukan, semua *dataset dummy* yang telah dipersiapkan dapat diproses dengan baik. Sudah dilakukan juga pengamatan pada setiap *file* PDF yang telah diunduh dan data yang dihasilkan di dalam tiap-tiap *file* PDF sudah sesuai dengan *dataset dummy* yang diprosesnya.

4.3 Pengujian *Contract Acceptance Test* (CAT)

4.3.1 Deskripsi Pengujian

Pengujian aplikasi *website* berdasarkan *Contract Acceptance Testing* (CAT) merupakan tahap yang signifikan dalam memastikan bahwa aplikasi telah memenuhi keperluan dan harapan para pengguna. Proses CAT dilakukan dengan melibatkan pengguna atau *stakeholder* yang mewakili pengguna akhir dari aplikasi tersebut. Dalam pengujian CAT, para pengguna akan secara langsung menguji aplikasi menggunakan situasi dan skenario penggunaan yang relevan dengan kebutuhan bisnis. Tujuan utama dari proses pengujian ini adalah untuk memverifikasi bahwa aplikasi beroperasi dengan baik, memiliki antarmuka yang *user-friendly*, dan memenuhi standar serta harapan pengguna yang sudah tertera pada kontrak sebelumnya. Dengan melibatkan pengguna dalam proses pengujian, dapat mengidentifikasi permasalahan, meningkatkan kualitas, dan memastikan bahwa aplikasi siap digunakan dan diterima secara positif oleh para pengguna akhir.

4.3.2 Prosedur Pengujian

Pengujian *Contract Acceptance Test* (CAT) dilakukan dengan responden yang akan menggunakan aplikasi ini nantinya. Dalam pengujian ini ada beberapa aspek yang dinilai, aspek – aspek tersebut dapat dilihat sebagai berikut.

- 1) *User Interface*
- 2) Proses Pada Sistem
- 3) Fungsionalitas Sistem

Berikut adalah prosedur pengujian *Contract Acceptance Test* (CAT) dalam pengujian aplikasi *website*:



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

1. Melibatkan responden yang merupakan nantinya akan menggunakan aplikasi ini.
2. Responden akan masuk ke dalam webiste dengan link yang telah diberikan
3. Setelah menggunakan sistem, responden akan diminta untuk memberikan penilaian mereka erdasarkan tiga aspek yang dinilai, yaitu *User Interface*, *Proses Pada Sistem*, dan *Fungsionalitas Sistem*
4. Untuk setiap aspek yang dinilai, pengguna akan diminta nutuk memberikan penilaiannya dalam salah satu dari lima kategori yang telah ditentukan; Sangat Setuju (SS), Setuju (S), Kurang Setuju (KS), Tidak Setuju (TS), dan Tidak Menjawa (TM).

4.3.3 Data Hasil Pengujian

Pengujian *Contract Acceptance Test* (CAT) dilakukan dengan responden sebanyak 3 orang. Bentuk *Contract Acceptance Test* (CAT) dikembangkan berdasarkan teori CAT pada BAB II dengan pertanyaan yang disesuaikan dengan sistem yang dibuat. Berikut merupakan Tabel 4.6 yang berisikan kesimpulan berdasarkan hasil secara menyeluruh.

Tabel 4.24 Hasil *Contract Acceptance Test* (CAT)

No	Aspek	SS	S	KS	TS	TM
1	User Interface	14	7	-	-	-
2	Proses Pada Sistem	8	4	-	-	-
3	Fungsionalitas Sistem	237	-	-	-	-

4.3.4 Analisa Hasil Pengujian

Berdasarkan data hasil pengujian *Contract Acceptance Test* yang telah dilakukan, selanjutnya dihitung persentase dari rata-rata penilaian pada setiap aspek. Berikut Tabel 4.7 yang berisikan persentase penilaian dari seluruh responden pada setiap aspek.

Tabel 4.25 Persentase Penilaian Pada Setiap Aspek

No	Aspek	SS (%)	S (%)	KS (%)	TS (%)	TM (%)
1	User Interface	66.66%	33.33%	0%	0%	0%
2	Proses Pada Sistem	66.66%	33.33%	0%	0%	0%
3	Fungsionalitas Sistem	100%	0%	0%	0%	0%

Berikut merupakan analisis hasil pengujian *Contract Acceptance Test* (CAT) pada setiap aspek:

1. *User Interface*

Mayoritas pengguna setuju bahwa *user interface* yang dibuat sesuai dengan yang diharapkan, dengan 66.66% sangat setuju dan 33.33% setuju.

2. *Proses Pada Sistem*

Mayoritas pengguna setuju bahwa proses pada sistem yang dibuat sesuai dengan yang diharapkan, dengan 66.66% sangat setuju dan 33.33% setuju.

3. *Fungsionalitas Sistem*

Mayoritas pengguna setuju bahwa fungsionalitas sistem yang dibuat sesuai dengan yang diharapkan, dengan 100% sangat setuju.

Dari hasil ini, dapat disimpulkan bahwa pengguna menyatakan puas dengan sistem *customer churn prediction* yang telah dibuat. Mayoritas pengguna setuju bahwa *user interface*, proses pada sistem, dan fungsionalitas sistem sesuai dengan yang mereka harapkan dan dapat memenuhi kebutuhan disepakati dari setiap pengguna, juga sudah sesuai dengan kontrak yang telah dibuat.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta