



**PENGGUNAAN METODE *RANDOM FOREST* DAN  
*SUPPORT VECTOR MACHINE* DALAM KLASIFIKASI  
BENCANA ALAM BERDASARKAN DATA *TWITTER***

**LAPORAN SKRIPSI**

**DIANA ANGGRAINI**

**4817070521**

**PROGRAM STUDI TEKNIK INFORMATIKA  
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER  
POLITEKNIK NEGERI JAKARTA**

**2021**



**PENGGUNAAN METODE *RANDOM FOREST* DAN  
*SUPPORT VECTOR MACHINE* DALAM KLASIFIKASI  
BENCANA ALAM BERDASARKAN DATA *TWITTER***

**LAPORAN SKRIPSI**

**Dibuat untuk Melengkapi Syarat-Syarat yang Diperlukan untuk  
Memperoleh Diploma Empat Politeknik**

**DIANA ANGGRAINI**

**4817070521**

**PROGRAM STUDI TEKNIK INFORMATIKA  
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER  
POLITEKNIK NEGERI JAKARTA**

**2021**



## HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Diana Anggraini

NIM : 4817070521

Tanggal : 29 Juli 2021

Tanda Tangan :

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 dari Jurusan TIK Politeknik Negeri Jakarta



© Hak Cipta milik Jurusan TIK 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 dari Jurusan TIK Politeknik Negeri Jakarta

## LEMBAR PENGESAHAN

Skripsi diajukan oleh :

Nama : Diana Anggraini  
NIM : 4817070521  
Program Studi : Teknik Informatika  
Judul Skripsi : Penggunaan Metode *Random Forest* dan Support *Vector Machine (SVM)* dalam Klasifikasi Bencana Alam Berdasarkan Data *Twitter*

Telah diuji oleh tim penguji dalam Sidang Skripsi pada hari Kamis, Tanggal 05, Bulan Agustus, Tahun 2021 dan dinyatakan **LULUS**.

Disahkan oleh

Pembimbing I : Mera Kartika Delimayanti, S.Si, M.T., Ph.D. (  )  
Penguji I : Euis Oktavianti, S.Si., M.Ti. (  )  
Penguji II : Dewi Kurniawati, S.S., M.Pd. (  )  
Penguji III : Rizki Elisa Nalawati, S.T., M.T. (  )

Mengetahui :

Jurusan Teknik Informatika dan Komputer

Ketua



Mauldy Laya, S.Kom., M.Kom.

NIP. 197802112009121003



**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 dari Jurusan TIK Politeknik Negeri Jakarta

## KATA PENGANTAR

Puji Syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan laporan skripsi ini. Penulisan laporan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat wajib untuk mencapai gelar Diploma Empat Politeknik Negeri Jakarta. Penulis menyadari bahwa penyusunan laporan skripsi ini tidak terlepas dari bantuan dan bimbingan dari berbagai pihak yang telah memberikan masukan kepada penulis. Oleh karena itu, penulis mengucapkan terima kasih kepada :

- a. Allah SWT, yang telah memberikan penulis rizki berupa kesehatan dan akal sehat yang sangatlah berharga bagi penulis sehingga laporan ini dapat terselesaikan dengan baik.
- b. Ibu Mera Kartika Delimayanti, S.Si, M.T., Ph.D. selaku dosen pembimbing yang telah menyediakan waktunya dalam mengarahkan dan membimbing penulis dalam penyusunan skripsi ini.
- c. Orang tua dan keluarga penulis yang telah memberikan bantuan dukungan secara moral maupun materil.
- d. Sahabat dan rekan sejawat yang telah banyak membantu dan mendukung penulis dalam menyelesaikan laporan skripsi ini.

Akhir kata, penulis berharap Allah SWT dapat membantu dan mempermudah penulis untuk dapat menyelesaikan hingga akhir dan semoga Allah SWT membalas kebaikan semua pihak yang telah membantu penulis dengan pahala yang berlipat ganda. Semoga laporan skripsi ini dapat bermanfaat bagi orang banyak.

Bekasi, 29 Juli 2021

Diana Anggraini



## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Politeknik Negeri Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Diana Anggraini  
NIM : 4817070521  
Program Studi : Teknik Informatika  
Jurusan : Teknik Informatika dan Komputer  
Jenis karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Politeknik Negeri Jakarta **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul :

“Penggunaan Metode Random Forest dan Support Vector Machine Dalam Klasifikasi Bencana Alam Berdasarkan Data Twitter”

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Politeknik Negeri Jakarta berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bekasi

Pada tanggal : 29 Juli 2021

Yang menyatakan

( Diana Anggraini )

### 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 dari Jurusan TIK Politeknik Negeri Jakarta



## Penggunaan Metode Random Forest dan Support Vector Machine (SVM) dalam Klasifikasi Bencana Alam Berdasarkan Data Twitter

### Abstrak

Twitter merupakan salah satu media sosial yang cukup populer saat ini. Karna hampir setiap orang menggunakan Twitter dalam kehidupan sehari-hari. Twitter memiliki fitur-fitur yang cukup sederhana tapi mampu membuat pengguna aktif memanfaatkannya dengan jenis layanan yang bersifat real-time. Pengguna meramaikan media sosial twitter dengan memberikan informasi (twit) tentang kejadian yang terjadi yang terjadi disekitarnya. Misalnya ketika terjadi bencana alam seperti, banjir, gempa dan kebakaran hutan. Pada tugas akhir ini akan dilakukan klasifikasi kejadian bencana alam (banjir, gempa dan kebakaran hutan) berdasarkan data twit yang dilakukan oleh pengguna yang menjadi saksi mata pada kejadian tersebut. Metode yang akan diimplementasikan adalah Random Forest dan Support Vector Machine (SVM). Nilai akurasi dari model klasifikasi yang paling tinggi akan dijadikan acuan sebagai metode klasifikasi yang terbaik dalam kasus ini. Selain itu, informasi bencana berupa titik kejadian akan ditampilkan ke sistem sebagai konsumsi user.

**Kata kunci :** Bencana alam, Random Forest, Support Vector, Machine (SVM), Twitter

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 kritis atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS .....	i
LEMBAR PENGESAHAN .....	ii
KATA PENGANTAR .....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iv
Abstrak .....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR .....	x
BAB I .....	1
PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Perumusan Masalah .....	2
1.3 Batasan Masalah .....	3
1.4 Tujuan dan Manfaat .....	3
1.4.1 Tujuan .....	3
1.4.2 Manfaat .....	3
1.5 Metode Pengumpulan Data .....	4
1.5.1 Scraping .....	4
1.5.2 <i>Preprocessing</i> .....	4
1.6 Metode Pelaksanaan .....	4
BAB II.....	6
TINJAUAN PUSTAKA .....	6
2.1 Penelitian Sebelumnya .....	6
2.2 TF-IDF.....	6
2.3 Metode Agile .....	8
2.4 Text Mining .....	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 mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta





## © Hak Cipta milik Jurusan TIK 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 dari Jurusan TIK Politeknik Negeri Jakarta

2.5	Bencana Alam.....	9
2.6	Scraping .....	9
2.7	Twitter .....	10
2.8	Dataset .....	10
2.9	<i>Preprocessing</i> .....	10
2.10	Python.....	10
2.12	Metode Support Vector Machine .....	11
2.13	Confusion Matrix.....	12
BAB III	.....	13
PERENCANAAN DAN REALISASI ATAU RANCANG BANGUN.....		13
3.1	Perancangan Program Aplikasi .....	13
3.1.1	Deskripsi Program Aplikasi.....	13
3.1.2	Cara Kerja Sistem.....	13
3.1.3	Rancangan Program Aplikasi .....	19
3.2	Realisasi Program .....	21
BAB IV	.....	40
PEMBAHASAN .....		40
4.1	Pengujian .....	40
4.2	Deskripsi Pengujian.....	40
4.3	Prosedur Pengujian .....	40
4.4	Data Hasil Pengujian .....	41
4.5	Analisis Data.....	43
BAB V	.....	45
PENUTUP.....		45
5.1	Kesimpulan.....	45
5.1	Saran .....	45
DAFTAR PUSTAKA .....		46



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Daftar Riwayat Penulis ..... 48



**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 dari Jurusan TIK 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 dari Jurusan TIK Politeknik Negeri Jakarta

**DAFTAR TABEL**

Tabel 1 Confusion Matrix .....	12
Tabel 2 Contoh <i>Preprocessing</i> .....	16
Tabel 3 Pengujian <i>K-Fold Cross Validation</i> .....	35
Tabel 4 Parameter Model Klasifikasi <i>Random Forest</i> .....	40
Tabel 5 Parameter Model Klasifikasi <i>Support Vector Machine</i> .....	41
Tabel 6 Peramalan Label Bencana.....	41
Tabel 7 Perbandingan Akurasi <i>Random Forest</i> .....	42
Tabel 8 Perbandingan Akurasi <i>Support Vector Machine</i> OVA.....	42
Tabel 9 Perbandingan Akurasi <i>Support Vector Machine</i> OVO.....	43





## DAFTAR GAMBAR

Gambar 2.1 Metode Agile.....	8
Gambar 3.1 <i>Flowchart</i> Sistem Keseluruhan .....	14
Gambar 3. 2 <i>Flowchart Preprocessing</i> .....	15
Gambar 3.3 Pemodelan .....	18
Gambar 3.4 Usecase Model .....	20
Gambar 3.5 <i>Scraping</i> .....	22
Gambar 3.6 Dataset.....	22
Gambar 3.7 <i>Preprocessing</i> – Dataset Bersih .....	23
Gambar 3.8 <i>Preprocessing</i> – Import Library.....	23
Gambar 3.9 <i>Preprocessing</i> – Baca Dataset.....	24
Gambar 3.10 <i>Preprocessing</i> - Cleansing .....	24
Gambar 3.11 <i>Preprocessing</i> – Case Folding.....	25
Gambar 3.12 <i>Preprocessing</i> - Tokenizing .....	25
Gambar 3.13 <i>Preprocessing</i> - Stemming.....	25
Gambar 3.14 <i>Preprocessing</i> - Stopword.....	26
Gambar 3.15 <i>Preprocessing</i> - Rejoin.....	26
Gambar 3. 16 <i>Preprocessing</i> - Mengubah Format Tanggal.....	26
Gambar 3.17 <i>Preprocessing</i> – Simpan ke Database .....	27
Gambar 3.18 Modelling – Koneksi ke Database .....	27
Gambar 3.19 Modelling – Convert Data.....	28
Gambar 3.20 Modelling – TF-IDF.....	28
Gambar 3.21 Modelling – K-Fold Cross Validation.....	29
Gambar 3.22 Modelling – Klasifikasi Random Forest .....	30
Gambar 3. 23 Modelling - Visualisasi Random Forest.....	30
Gambar 3.24 Modelling - Visualisasi Random Forest.....	31
Gambar 3.25 Modelling - Visualisasi Random Forest.....	31
Gambar 3.26 Modelling - Visualisasi Random Forest.....	32
Gambar 3.27 Visualisasi <i>Random Forest</i> - Data Keseluruhan .....	32
Gambar 3. 28 Visualisasi Random Forest - Data Balance .....	33
Gambar 3.29 Visualisasi Random Forest - Confusion Matrix.....	34
Gambar 3. 30 Visualisasi Random Forest – Tree Estimator.....	34

### 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 dari Jurusan TIK Politeknik Negeri Jakarta



## © Hak Cipta milik Jurusan TIK 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 dari Jurusan TIK Politeknik Negeri Jakarta

Gambar 3.31 Modelling – Klasifikasi Support Vector Machine OVA.....	35
Gambar 3.32 Modelling – Klasifikasi Support Vector Machine OVO.....	35
Gambar 3.33 Testing - Confusion Matrix Random Forest .....	36
Gambar 3.34 Testing - Confusion Matrix Support Vector Machine OVA.....	37
Gambar 3.35 Testing – Confusion Matrix Support Vector Machine OVO .....	38
Gambar 3.36 Pengaturan <i>Real - Time</i> .....	39
Gambar 3. 37 Identifikasi Label Baru.....	39





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 dari Jurusan TIK Politeknik Negeri Jakarta

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang Masalah

Perkembangan teknologi terutama pada media digital mendorong manusia untuk dapat berkomunikasi dan membagikan kesehariannya pada berbagai media digital yang dapat diakses secara online. Media digital ini dikenal dengan sebutan media sosial. Salah satu media sosial yang cukup akrab dengan berbagai kalangan adalah *Twitter*. *Twitter* adalah sosial media yang memberikan layanan pada penggunanya untuk berbagi hal-hal yang sedang terjadi saat itu juga (Dwi Lingga P., 2017). Media sosial ini digunakan dengan cara menuliskan status berupa note dengan jumlah karakter yang dibatasi yang disebut dengan *twit*. Selain membuat *twit* dalam bentuk tulisan, *twit* juga dapat memuat foto dan video. Fitur-fitur yang disediakan oleh *Twitter* juga akan dimuat secara *real-time*. Hal ini merupakan salah satu pendorong yang membuat pengguna dapat berlama-lama untuk menggunakan media sosial ini.

Dengan menggunakan media sosial *Twitter*, pengguna biasanya membagikan apapun tentang kesehariannya. Tidak terkecuali pada bencana alam yang terjadi disekitarnya. Pengguna *Twitter* termasuk yang cukup aktif dalam membagikan informasi terkait perkembangan bencana alam yang terjadi disekitarnya. Seperti ketika terjadi banjir, gempa dan kebakaran hutan. Hal ini cukup membantu pengguna lain untuk mengetahui informasi terbaru terkait bencana yang terjadi. Selain itu pengguna lainnya juga dapat mengantisipasi tentang prediksi lokasi-lokasi yang terdampak.

Akan tetapi, tentu dibutuhkan kevalidan data untuk dapat dijadikan informasi yang kuat untuk dikonsumsi oleh user. Dalam hal ini, penghimpunan data dilakukan terhadap 3 akun *Twitter* terkait kebencanaan, yaitu BMKG, BNPB dan SAR Nasional. Pada penelitian sejenis dengan menggunakan algoritma *Naive Bayes* yang dilakukan (Hamidah, 2019) mengatakan BMKG merupakan sebuah lembaga instansi pemerintahan yang bertugas dibidang Meteorologi, Klimatologi, Kualitas Udara dan Geofisika sesuai perundang - undangan yang ada. Akun



## © Hak Cipta milik Jurusan TIK 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 dari Jurusan TIK Politeknik Negeri Jakarta

Twitter BMKG berisi informasi tentang perkiraan cuaca terkini serta bencana alam yang ada di Indonesia. Sedangkan, BNPB merupakan lembaga instansi pemerintahan yang berfungsi untuk merumuskan dan menetapkan kebijakan penanggulangan bencana dan penanganan pengungsi dengan bertindak cepat dan tepat serta dapat mengkoordinasikan pelaksanaan kegiatan penanggulangan bencana secara terencana, terpadu, dan menyeluruh (Hamidah, 2019). Sementara itu, SAR Nasional merupakan lembaga pemerintahan yang bertugas melaksanakan tugas pemerintahan di bidang pencarian dan pertolongan.

Kebutuhan informasi tentu saja tidak hanya diperlukan para pengguna *Twitter*. Masyarakat secara umum tentu saja juga memerlukan informasi terkait bencana alam yang sedang terjadi tersebut. Sehingga masyarakat secara umum juga dapat mengantisipasi dirinya dengan berbagai cara yang dapat mereka lakukan berdasarkan informasi yang diterima.

Berdasarkan permasalahan tersebut, dilakukanlah perancangan dan pembangunan sistem penerapan kecerdasan buatan dalam memprediksi bencana alam dengan pemanfaatan *social network sensor*. Sistem ini akan menampilkan lokasi-lokasi yang sedang dilanda bencana banjir, gempa maupun kebakaran hutan. Selain itu, sistem ini juga dapat menampilkan akurasi dari data yang ditampilkan. Dimana data tersebut akan dihimpun dari Twitter dengan menggunakan *scrapping*. Setelah itu, data tersebut akan diklasifikasikan dengan menggunakan metode *Random Forest* dan *Support Vector Machine (SVM)*. Nilai akurasi yang paling tinggi akan ditampilkan dalam sistem yang kemudian dapat dijadikan acuan sebagai keakuratan dalam memprediksi bencana alam.

### 1.2 Perumusan Masalah

Adapun perumusan masalah dalam perancangan dan pembuatan sistem sebagai berikut.

1. Bagaimana pengimplementasian *scrapping* terhadap pengumpulan data dari *Twitter*.
2. Bagaimana pengimplementasian metode klasifikasi dalam menampilkan keakuratan informasi tentang bencana alam yang terjadi.



**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 dari Jurusan TIK Politeknik Negeri Jakarta

### 1.3 Batasan Masalah

Berikut ini merupakan batasan masalah terkait permasalahan yang telah diuraikan.

1. Bencana alam yang dibahas adalah banjir, gempa dan kebakaran hutan.
2. Data dihimpun dari akun twitter BMKG, BNPB dan SAR Indonesia
3. Sumber penghimpunan data diperoleh dari *Twitter*.
4. Klasifikasi menggunakan metode *Random Forest* dan *Support Vector Machine (SVM)*.

### 1.4 Tujuan dan Manfaat

Berikut ini merupakan tujuan dan manfaat dari perancangan dan pembangunan sitem terkait permasalahan yang telah di uraikan.

#### 1.4.1 Tujuan

Berikut ini merupakan beberapa tujuan dari perancangan dan pembangunan sistem terkait permasalahan yang telah diuraikan.

1. Implementasi *scraping* terhadap penghimpunan data dari *Twitter*.
2. Menentukan metode klasifikasi yang akan digunakan untuk menampilkan keakuratan informasi bencana alam yang terjadi.

#### 1.4.2 Manfaat

Berikut ini merupakan beberapa manfaat dari perancangan dan pembangunan sistem terkait permasalahan yang telah diuraikan.

1. Memperoleh data tentang bencana terkait setelah dilakukan penghimpunan data menggunakan *scraping*.
2. Menampilkan perbandingan akurasi berita bencana terkait melalui beberapa metode yang di gunakan





**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 dari Jurusan TIK Politeknik Negeri Jakarta

## 1.5 Metode Pengumpulan Data

### 1.5.1 Scraping

Adapun metode pengumpulan data untuk kebutuhan penelitian ini menggunakan *scraping* pada data *twitter*. Scraping adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet (Budiman and Widiksono, 2018).

### 1.5.2 Preprocessing

Text *Preprocessing* adalah tahap awal dari text mining. Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi knowledge discovery sistem text mining. Tindakan yang dilakukan pada tahap ini adalah *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil dan *Tokenizing* yang berfungsi sebagai penguraian deskripsi yang semula berupa kalimat menjadi kata-kata lalu menghilangkan delimiter-delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka (Kirana *et al.*, 2019).

## 1.6 Metode Pelaksanaan

Penulisan tugas akhir ini menggunakan salah satu metode dari bagian *SDLC* (*System Development Life Cycle*). Yaitu menggunakan pendekatan *Agile*. *Agile* sendiri merupakan pengembangan perangkat lunak yang memungkinkan perubahan sesuai kebutuhan dan mengutamakan kecepatan dalam pengembangannya. Berikut merupakan tahapan dari metode *agile* :

- a) *Requirement*, tahapan yang dilakukan dengan merancang konsep aplikasi yang akan dibuat untuk mengetahui *user requirement* dan alur pengembangan aplikasinya.
- b) *Design*, tahapan yang dilakukan dengan mendesain konsep aplikasi sesuai *user requirement* menjadi alur dan *usecase* untuk membantu proses pengembangan dalam pembuatan aplikasi.
- c) *Implementation*, tahapan ini merealisasikan desain dan rancangan yang dibuat melalui *coding* dan imolementasi algoritma dengan menggunakan *data Preprocessing* dan *modelling* sesuai kebutuhan fitur *user requirement*.

- d) *Testing*, tahapan ini melakukan pengujian terhadap fitur-fitur yang telah dibuat guna mencari *bug* dan menemukan *error*. Pada tahapan ini juga akan dilakukan proses evaluasi untuk mengevaluasi *error* terhadap model yang telah dibuat.
- e) *Review*, tahapan ini dilakukan melihat dan menilai sistem yang dibuat.
- (Hapizhudin, 2020).



## © Hak Cipta milik Jurusan TIK 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 dari Jurusan TIK 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 mengumukkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

1. Penelitian ini menghasilkan model klasifikasi bencana alam yang dibuat dengan algoritma Random Forest dan Support Vector Machine.
2. Penelitian ini dapat digunakan untuk melakukan peramalan pada label jenis bencana
3. Pada implementasi dan pengujian yang dilakukan, nilai akurasi tertinggi konstan terdapat pada *Support Vector Machine* OVA. Dimana nilai akurasi yang didapat *Support Vector Machine* OVA pada implementasi sebesar 98.69%, sedangkan *Random Forest* dengan akurasi 94.44% dan *Support Vector Machine* OVO dengan akurasi 96.73%. Selain itu, *Support Vector Machine* OVA juga menampilkan akurasi tertinggi pada pengujian yang dibuktikan pada perbandingan tabel 7, tabel 8 dan tabel 9.
4. Berdasarkan review dan pengujian tersebut, penelitian ini memperoleh bahwa model klasifikasi *Support Vector Machine* OVA merupakan metode klasifikasi yang terbaik.

#### 5.1 Saran

1. Pastikan setiap tahapan dari *Preprocessing* dapat berkerja dengan baik, sehingga data benar-benar bersih
2. Pastikan data yang digunakan tidak overfitting, sehingga nilai akurasi lebih optimal.



**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 dari Jurusan TIK Politeknik Negeri Jakarta

**DAFTAR PUSTAKA**

- Ankit and Saleena, N. (2018) 'An Ensemble Classification System for Twitter Sentiment Analysis', *Procedia Computer Science*, 132(Iccids), pp. 937–946. doi: 10.1016/j.procs.2018.05.109.
- Budiman, A. A. and Widiksono, S. (2018) 'Aplikasi Pengolahan Data Untuk Menganalisa Penggunaan Hashtag Pada Twitter', 8(2). Available at: <http://www.twitter.com>.
- Delimayanti, M. K., Sari, R., Laya, M., Faisal, M. R. and Pahrul (2020) 'Pengaruh Teknik Klasifikasi Pada Pesan Bencana Banjir Di Twitter Dengan Metode Multiclass-SVM', 8(1), p. 4.
- Delimayanti, M. K., Sari, R., Laya, M., Faisal, M. R., Pahrul, *et al.* (2020) 'The effect of pre-processing on the classification of twitter's flood disaster messages using support vector machine algorithm', *Proceedings of ICAE 2020 - 3rd International Conference on Applied Engineering*. doi: 10.1109/ICAE50557.2020.9350387.
- Dwi Lingga P., R. (2017) 'Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision Tree, Random Forest, dan SVM', *Jurnal Teknik ITS*, 6(1). doi: 10.12962/j23373539.v6i1.22037.
- Hafidz, N. and Liliana, D. Y. (2021) 'Klasifikasi Sentimen pada Twitter Terhadap WHO', 1(10), pp. 3–4.
- Hamidah, N. N. (2019) 'Sistem Informasi Pemetaan Bencana Daerah Wilayah Dengan Data Twitter', 2009.
- Handayani, A. S. *et al.* (2020) 'i ISSN . 2620-6900 ( Online ) 2620-6897 ( Cetak )', 3(1).
- Hapizhudin, A. F. (2020) 'Implementasi sentiment analysis cyberbullying pada twitter dengan algoritma support vector machine laporan skripsi'.



## © Hak Cipta milik Jurusan TIK 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 dari Jurusan TIK Politeknik Negeri Jakarta

Kirana, M. C. *et al.* (2019) ‘Visualisasi Kualitas Penyebaran Informasi Gempa Bumi di Indonesia Menggunakan Twitter’, *Journal of Applied Informatics and Computing*, (May), pp. 23–32. doi: 10.30871/jaic.v0i0.1246.

Luqyana, W. A., Cholissodin, I. and Perdana, R. S. (2018) ‘Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(11), pp. 4704–4713.

Patel, R. and Passi, K. (2020) ‘Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning’, *IoT*, 1(2), pp. 218–239. doi: 10.3390/iot1020014.

Suliztia, M. L. (2020) ‘Penerapan Analisis Random Forest Pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask’.



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 dari Jurusan TIK Politeknik Negeri Jakarta

## Daftar Riwayat Penulis

### Diana Anggraini



Lahir di Tanjungpinang, 24 Mei 1999. Lulus dari SDN 004 Toapaya pada tahun 2011, SMPN 17 Bintan pada tahun 2014 dan SMAN 1 Toapaya pada tahun 2017. Saat ini sedang menempuh pendidikan Diploma IV Program Studi Teknik Informatika Jurusan Teknik Informatika dan Komputer di Politeknik Negeri Jakarta.



**POLITEKNIK  
NEGERI  
JAKARTA**