



© 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 mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**IMPLEMENTASI SISTEM PENCARIAN GAMBAR
BERBASIS TEKS BAHASA INDONESIA DENGAN FINE-
TUNED CLIP TEXT ENCODER DAN VISION
TRANSFORMER IMAGE ENCODER**

SKRIPSI

**POLITEKNIK
NEGERI
JAKARTA**

**Muhammad Nur Irfan
2207411056**

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN
KOMPUTER**

POLITEKNIK NEGERI JAKARTA

2026



© 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 mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**IMPLEMENTASI SISTEM PENCARIAN GAMBAR
BERBASIS TEKS BAHASA INDONESIA DENGAN FINE-
TUNED CLIP TEXT ENCODER DAN VISION
TRANSFORMER IMAGE ENCODER**

SKRIPSI

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

Muhammad Nur Irfan

2207411056

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

2026



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

SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertanda tangan di bawah ini:

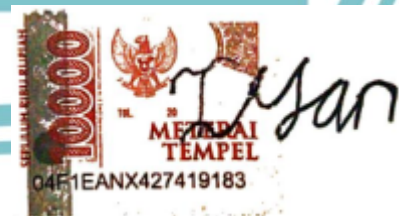
Nama : Muhammad Nur Irfan
NIM : 2207411056
Jurusan/Program Studi : T.Informatika dan Komputer / Teknik Informatika
Judul skripsi : IMPLEMENTASI SISTEM PENCARIAN GAMBAR BERBASIS TEKS BAHASA INDONESIA DENGAN FINE-TUNED CLIP TEXT ENCODER DAN VISION TRANSFORMER IMAGE ENCODER

Menyatakan dengan sebenarnya bahwa skripsi ini benar-benar merupakan hasil karya saya sendiri, bebas dari peniruan terhadap karya dari orang lain. Kutipan pendapat dan tulisan orang lain ditunjuk sesuai dengan cara-cara penulisan karya ilmiah yang berlaku.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa dalam skripsi ini terkandung cirri-ciri plagiat dan bentuk-bentuk peniruan lain yang dianggap melanggar peraturan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Jakarta, 9 Juni 2026

Yang membuat pernyataan



Muhammad Nur Irfan

NIM 2207411056



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

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Politeknik Negeri Jakarta, saya bertanda tangan dibawah ini :

Nama : Muhammad Nur Irfan
NIM : 2207411056
Jurusan/Program Studi : T.Informatika dan Komputer / Teknik Informatika

Demi pengembangan ilmu pengetahuan , menyetujui untuk memberikan kepada Politeknik Negeri Jakarta Hak Bebas Royalti Non-Eksklusif atas karya ilmiah sayayang berjudul :

IMPLEMENTASI SISTEM Pencarian Gambar Berbasis Teks
Bahasa Indonesia dengan Fine-Tuned CLIP Text Encoder
dan Vision Transformer Image Encoder

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksklusif ini Politeknik Negeri Jakarta Berhak menyimpan, mengalihmediakan/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 9 Juni 2026

Yang membuat pernyataan




Muhammad Nur Irfan

2207411056



© 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 mengumukan 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 : Muhammad Nur Irfan
NIM : 2207411056
Program Studi : Teknik Informatika
Judul Skripsi : Implementasi Sistem Pencarian Gambar Berbasis Teks Bahasa Indonesia Dengan *Fine-Tuned CLIP Text Encoder* dan *Vision Transformer Image Encoder*

Telah diuji oleh tim penguji dalam Sidang Skripsi pada hari Senin, Tanggal 22, Bulan Juni, Tahun 2026 dan dinyatakan **LULUS**.

Disahkan oleh

Pembimbing I : Euis Oktavianti, S.Si., M.T.I

(.....)

Penguji I : Dr. Dewi Yanti Liliana, S.Kom., M.Kom.

(.....)

Penguji II : Dr. Indra Hermawan, M.Kom.

(.....)

Penguji III : Anggi Mardiyono, S.Kom., M.Kom.

(.....)

Mengetahui:

Jurusan Teknik Informatika dan Komputer

Ketua



Dr. Anita Hidayati, S.Kom., M.Kom.

NIP. 197908032003122003



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

Assalamualaikum Warahmatullahi Wabarakatuh, Puji Syukur saya panjatkan kepada Tuhan Yang Maha Esa, yang telah memberikan karunia dan rahmat-Nya kepada penulis, sehingga penulis dapat menyelesaikan skripsi yang berjudul “IMPLEMENTASI SISTEM Pencarian Gambar Berbasis Teks Bahasa Indonesia dengan Fine-Tuned CLIP Text Encoder dan Vision Transformer Image Encoder” sebagai salah satu syarat untuk mencapai gelar Sarjana Terapan Program Studi Teknik Informatika Politeknik Negeri Jakarta.

Penulis sangat menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, sangatlah sulit bagi penulis untuk menyelesaikan skripsi ini. Dengan demikian, penulis ingin mengucapkan terima kasih kepada pihak-pihak yang terlibat. Secara khusus, penulis menyampaikan ucapan terimakasih kepada:

1. Ibu Dr. Anita Hidayati, S. Kom., M. Kom., selaku Ketua Jurusan Teknik Informatika dan Komputer.
2. Ibu Euis Oktavianti, S.Si., M.TI., selaku Kepala Program Studi Teknik Informatika sekaligus Dosen Pembimbing yang sudah bersedia meluangkan waktu untuk membimbing, mengarahkan dan menyemangati dalam proses penyelesaian skripsi ini.
3. Seluruh Bapak/Ibu dosen yang sudah mendidik penulis sehingga menjadi pribadi yang lebih baik
4. Kedua orang tua penulis yang senantiasa mendukung, memberikan tempat cerita, memberikan doa dan semangat serta kasih sayang tanpa henti kepada penulis
5. Azzamahdy Ahmad Rafiqri selaku saudara penulis yang memberikan semangat, nasihat dan tempat untuk berdiskusi kepada penulis
6. Sahabat Prajurit selaku teman seperjuangan dan teman teman lainnya yang memberikan dukungan, solusi dan menemani penulis dalam penyelesaian skripsi ini



© 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 mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Akhir kata, penulis berharap skripsi yang ditulis ini dapat memberikan manfaat kepada berbagai pihak. Penulis menyadari bahwa skripsi ini jauh dari kata

sempurna, oleh karena itu, penulis mengharapkan segala bentuk kritik, saran, dan masukan yang membangun yang dapat memperbaiki serta menyempurnakan skripsi ini.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Jakarta, 9 Juni 2026

Muhammad Nur Irfan





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

IMPLEMENTASI SISTEM Pencarian Gambar Berbasis Teks Bahasa Indonesia dengan Fine-Tuned CLIP Text Encoder dan Vision Transformer Image Encoder

ABSTRAK

Sistem pencarian gambar konvensional umumnya masih bergantung pada metadata manual yang kurang mampu menangkap makna semantik visual. Pendekatan cross-modal retrieval menggunakan model CLIP menawarkan solusi, namun performanya menurun pada bahasa non-Inggris seperti Bahasa Indonesia. Penelitian ini melakukan studi komparatif terkontrol terhadap tiga konfigurasi text encoder pada arsitektur CLIP (ViT-B/32), yaitu CLIP standar (bawaan), CLIP dengan IndoBERT, dan CLIP dengan XLM-RoBERTa untuk mengevaluasi efektivitas contrastive learning berbasis Bahasa Indonesia menggunakan dataset Flickr8K dan Flickr30K Indonesia. Hasil evaluasi metrik retrieval dua arah menunjukkan bahwa CLIP standar secara konsisten mencatatkan performa tertinggi pada kedua dataset (rata-rata Recall@10 mencapai 90,52% pada Flickr8K), diikuti oleh CLIP XLM-RoBERTa, sementara CLIP IndoBERT menghasilkan performa terendah. Temuan ini membuktikan bahwa keselarasan ruang embedding (embedding space alignment) sejak fase pre-training jauh lebih mendominasi performa dibandingkan faktor spesialisasi linguistik text encoder monolingual dalam kondisi data terbatas. Model terbaik (CLIP standar) berhasil diimplementasikan ke dalam sistem Smart Gallery berbasis web menggunakan FastAPI dan NextJS. Pengujian usability menggunakan System Usability Scale (SUS) menghasilkan skor rata-rata 91,36 (kategori Excellent, Grade A), menegaskan bahwa aplikasi sangat layak dan mudah dioperasikan oleh pengguna.

Kata Kunci: Bahasa Indonesia, Contrastive Learning, Cross-Modal Retrieval, CLIP, IndoBERT, Pencarian Gambar, Smart Gallery, XLM-RoBERTa.



DAFTAR ISI

KATA PENGANTAR.....	iii
ABSTRAK	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	x
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan dan Manfaat	3
1.4.1 Tujuan	3
1.4.2 Manfaat	3
1.5 Sistematika Penulisan	3
BAB 2 TINJAUAN PUSTAKA.....	5
2.1 Smart Gallery	5
2.1.1 Metadata.....	5
2.2 Cross-Modal Retrieval	6
2.2.1 <i>Embedding</i> pada Cross-Modal Retrieval	6
2.3 Contrastive Learning.....	6
2.4 Arsitektur Model	7
2.4.1 Transformer dan self attention	7
2.4.2 Bidirectional Encoder Representations from Transformers (BERT).....	8
2.4.3 Vision Transformer (ViT-B/32).....	9
2.4.4 Contrastive Language Image Pre-trained (CLIP)	11
2.4.5 IndoBERT	12
2.4.6 XLM-RoBERTa.....	13
2.5. Cross Industry Standard Process - Data Mining (CRISP-DM).....	14
2.5.1 Business Understanding.....	14
2.5.2 Data Understanding.....	14
2.5.3 Data Preparation.....	14
2.5.4 Modelling	15
2.5.5 Evaluation	15
2.5.6 Deployment	17
2.6 Tokenization.....	17
2.7 Software Development Life-cycle (SDLC)	18
2.7.1 Planning	18
2.7.2 Analysis.....	18
2.7.3 Design	19
2.7.4 Implementation	19
2.7.5 Testing.....	20
2.8 Penelitian Terkait	20
BAB 3 PEMBAHASAN	22
3.1 Rancangan Penelitian	22
3.1.2 Tahapan Penelitian	23
3.2 Sumber Data.....	26

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 :

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

3.3 Teknik Pengumpulan Data.....	27
3.4 Teknik Analisis Data.....	28
BAB 4 HASIL DAN PEMBAHASAN	29
4.1 Identifikasi Kebutuhan	29
4.1.1 Kebutuhan Pengembangan Model	29
B. Kebutuhan Model.....	30
4.1.2 Kebutuhan Website	30
4.1.3 Kebutuhan Perangkat Keras.....	31
4.1.4 Kebutuhan Perangkat Lunak.....	32
4.2 Perancangan Sistem	34
4.2.1 Perancangan Model Deep Learning.....	34
4.2.2 Perancangan Web.....	43
4.3 Implementasi Sistem	52
4.3.1 Persiapan Dataset	52
4.3.2 Implementasi Model.....	56
4.3.3 Implementasi Model pada Aplikasi Website	69
4.4 Pengujian Sistem.....	76
4.4.1 Deskripsi Pengujian	76
4.4.2 Prosedur Pengujian	77
4.4.3 Data Hasil Pengujian.....	79
4.4.4 Analisis Data	85
BAB 5 PENUTUP.....	87
5.1 Kesimpulan	87
5.2 Saran.....	88
DAFTAR PUSTAKA.....	89
LAMPIRAN.....	96

**POLITEKNIK
NEGERI
JAKARTA**

DAFTAR GAMBAR

Gambar 2.1	Arsitektur Transformer.....	8
Gambar 2.2	Arsitektur BERT	9
Gambar 2.3	Arsitektur Vision Transformer	10
Gambar 2.4	Arsitektur CLIP	11
Gambar 2.3	Rumus <i>Cosine similarity</i>	15
Gambar 2.4	Rumus Recall@K.....	16
Gambar 2.5	Rumus MRR.....	16
Gambar 2.6	Siklus SDLC.....	18
Gambar 3.1	Rancangan Penelitian	23
Gambar 4.1	Alur Tahapan Pelatihan Model	34
Gambar 4.2	Alur Pelatihan Contrastive Learning.....	36
Gambar 4.3	Arsitektur Image Encoder ViT-B/32.....	37
Gambar 4.4	Arsitektur text encoder CLIP standar.....	39
Gambar 4.5	Arsitektur text encoder menggunakan IndoBERT	40
Gambar 4.6	Arsitektur text encoder menggunakan XLM-RoBERTa.....	41
Gambar 4.7	Use Case Diagram.....	43
Gambar 4.8	Activity Diagram Halaman Utama.....	44
Gambar 4.9	Activity Diagram Pencarian Berdasarkan Teks	45
Gambar 4.10	Activity Diagram Pencarian Berbasis Gambar	46
Gambar 4.11	Activity Diagram Halaman About Gallery	47
Gambar 4.12	Activity Diagram Pengubahan Jumlah Gambar.....	47
Gambar 4.14	Arsitektur Sistem.....	49
Gambar 4.15	Contoh Dataset Caption	53
Gambar 4.16	Kode Pemrosesan Data Gambar.....	54
Gambar 4.17	Tokenization Sesuai dengan Nama Model.....	55
Gambar 4.18	Kode Data Loader Dataset	56
Gambar 4.19	Grafik Pelatihan Model CLIP pada Dataset Flickr8K	58
Gambar 4.20	Grafik Pelatihan Model CLIP pada Dataset Flickr30K	59
Gambar 4.21	Grafik Pelatihan Model CLIP IndoBERT pada Dataset Flickr8K..	61
Gambar 4.22	Pelatihan Model CLIP IndoBERT pada Dataset Flickr30K	62
Gambar 4.23	Pelatihan Model CLIP XLM-RoBERTa pada Dataset Flickr8K....	64
Gambar 4.24	Pelatihan Model CLIP XLM-RoBERTa pada Dataset Flickr30K..	65
Gambar 4.25	Alur Implementasi Aplikasi Website	70
Gambar 4.26	Tampilan Halaman Utama	71
Gambar 4.27	Tampilan Halaman Hasil Pencarian Gambar dengan Teks.....	72
Gambar 4.28	Tampilan Halaman Pencarian dengan Gambar.....	73
Gambar 4.29	Tampilan Halaman Hasil Pencarian Gambar dengan Gambar.....	74
Gambar 4.30	Tampilan Halaman About Gallery	75

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 2.1 Penelitian Terkait	20
Tabel 3.1 Perbandingan Dataset.....	27
Tabel 4.1 Kebutuhan Dataset	29
Tabel 4.2 Kebutuhan Model.....	30
Tabel 4.3 Kebutuhan Fungsional	31
Tabel 4.4 Kebutuhan Non-Fungsional	31
Tabel 4.5 Kebutuhan Perangkat Keras Pengembangan Model.....	32
Tabel 4.6 Kebutuhan Perangkat Keras Pengembangan Website	32
Tabel 4.7 Kebutuhan Perangkat Lunak Pengembangan Model.....	32
Tabel 4.8 Kebutuhan Perangkat Lunak Pengembangan Website	33
Tabel 4.9 Komponen Struktur Database	51
Tabel 4.10 Contoh Dataset Gambar dan Caption.....	52
Tabel 4.11 Contoh Pemrosesan Dataset Gambar	54
Tabel 4.12 Parameter Model CLIP	57
Tabel 4.13 Parameter Model CLIP dengan IndoBERT	60
Tabel 4.14 Parameter Model CLIP dengan XLM-RoBERTA.....	63
Tabel 4.15 Hasil Evaluasi Pelatihan dengan Dataset Flickr8K Indonesia	66
Tabel 4.16 Hasil Evaluasi Pelatihan dengan Dataset Flickr30K Indonesia	67
Tabel 4.17 Prosedur Pengujian Black Box	78
Tabel 4.18 Pertanyaan SUS	79
Tabel 4.19 Hasil Image Retrieval berdasarkan Kueri Teks	80
Tabel 4.20 Hasil Pengujian Black Box	82
Tabel 4.21 Hasil Pengujian SUS.....	83
Tabel 4.22 Rumus Perhitungan SUS.....	83
Tabel 4.23 <i>Rating</i> nilai SUS	84
Tabel 4.24 Skor Perhitungan SUS	84

**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

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Bahasa Indonesia merupakan bahasa dengan jumlah penutur lebih dari 198 juta jiwa dan menjadi salah satu bahasa dengan penetrasi internet tertinggi di Asia Tenggara, dengan 185,3 juta pengguna internet aktif pada tahun 2024 (Kemp, 2024). Pertumbuhan pengguna digital ini mendorong peningkatan produksi konten visual yang diproduksi dan dikonsumsi, mulai dari galeri pribadi hingga platform berbagi media. Namun, sistem pencarian gambar yang umum digunakan saat ini masih bergantung pada metadata manual seperti nama berkas dan tag subjektif yang tidak mampu menangkap makna semantik dari konten visual (Aske & Giardinetti, 2023). Kesenjangan ini menyulitkan pengguna berbahasa Indonesia untuk menemukan gambar yang relevan menggunakan deskripsi teks dalam bahasa Indonesia.

Pendekatan cross-modal retrieval berbasis *deep learning*, khususnya CLIP (Radford et al., 2021), telah terbukti efektif dalam menghubungkan representasi gambar dan teks dalam satu ruang semantik melalui *contrastive learning*. Namun, Dataset pelatihan CLIP didominasi oleh deskripsi berbahasa Inggris, sehingga bahasa-bahasa yang kurang terwakili termasuk bahasa-bahasa Asia Tenggara mengalami kesenjangan performa dibandingkan Bahasa Inggris (Santos et al., 2023).

Beranjak dari kebutuhan akan sistem *cross modal retrieval* yang mampu memahami konteks Bahasa Indonesia secara lebih mendalam, penelitian ini melakukan studi komparatif terhadap tiga pendekatan text encoder, yaitu CLIP, CLIP dengan XLM-RoBERTa, dan CLIP dengan IndoBERT untuk mengevaluasi efektivitas model dengan pendekatan *contrastive learning*. Oleh karena itu, hasil penelitian ini diharapkan mampu memberikan landasan empiris tentang efektivitas komparatif text encoder untuk arsitektur CLIP dalam konteks cross-modal retrieval berbahasa Indonesia, sekaligus menghadirkan sistem pencarian gambar 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 mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

memungkinkan pengguna melakukan pencarian secara intuitif menggunakan teks Bahasa Indonesia sehari-hari tanpa bergantung pada tag atau nama file.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah

1. Bagaimana perbandingan performa konfigurasi text encoder CLIP, CLIP dengan XLM-RoBERTa , dan CLIP dengan IndoBERT dengan *contrastive learning* untuk tugas pencarian gambar dengan kueri berbahasa Indonesia?
2. Bagaimana mengimplementasikan model *text encoder* terbaik ke dalam sistem Smart Gallery untuk mendukung pencarian gambar multimodal berbahasa Indonesia?

1.3 Batasan Masalah

Dalam penelitian ini, batasan masalah diperlukan agar penelitian yang dilakukan tetap terarah dan sistematis sesuai dengan perumusan masalah. Berikut adalah beberapa batasan masalah dalam penelitian ini:

1. Penelitian ini menggunakan dua dataset: Flickr8K Indonesia sebagai eksperimen pendahuluan dan Flickr30K Indonesia sebagai dataset utama evaluasi.
2. Penelitian ini menggunakan arsitektur Vision Transformer ViT-B/32 sebagai image encoder yang dibekukan (frozen) pada seluruh konfigurasi model.
3. Penelitian ini membandingkan tiga konfigurasi *text encoder*: CLIP standar, CLIP dengan XLM-RoBERTa, dan CLIP dengan IndoBERT sebagai *text encoder*.
4. Seluruh konfigurasi model dilatih menggunakan metode *contrastive learning* dengan fungsi loss InfoNCE, *batch size*, *optimizer* dan jumlah *epoch* yang identik, serta *learning rate* yang disesuaikan per arsitektur.
5. Konfigurasi freeze layer pada *text encoder* disesuaikan per arsitektur sehingga jumlah *layer trainable* setara di angka 6 *layer* untuk setiap model BERT, meskipun proporsi *freeze* berbeda akibat perbedaan kedalaman arsitektur.



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 mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

6. Penelitian ini menggunakan metrik Recall@K dan MRR (*Mean Reciprocal Rank*) untuk mengukur efektivitas *cross-modal retrieval* secara kuantitatif.
7. Sistem Smart Gallery yang dikembangkan mendukung pencarian gambar berbasis kueri teks dan kueri gambar, tanpa fitur penambahan gambar oleh pengguna.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Berdasarkan latar belakang dan rumusan masalah, tujuan utama dari penelitian ini adalah membandingkan efektivitas konfigurasi *text encoder* CLIP, CLIP dengan XLM-RoBERTa, dan CLIP dengan IndoBERT dengan *contrastive learning* untuk tugas *cross-modal retrieval* berbahasa Indonesia dan mengimplementasi model ke dalam sistem pencarian gambar.

1.4.2 Manfaat

Adapun manfaat dari penelitian ini, dapat dijabarkan sebagai berikut:

1. Memudahkan proses pencarian gambar menggunakan bahasa alami tanpa ketergantungan pada *metadata*, *tag*, atau nama file manual.
2. Mempercepat proses pencarian gambar dengan memberikan pencarian gambar berdasarkan fitur visual gambar.

1.5 Sistematika Penulisan

Untuk mempermudah penulisan penelitian untuk penulis dan pembaca, penulisan disusun secara sistematis dengan struktur penulisan berikut ini:

BAB I PENDAHULUAN

Bab ini berisi uraian mengenai latar belakang penelitian, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan yang menjadi gambaran umum arah dan ruang lingkup penelitian.



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 II LANDASAN TEORI

Bab ini memuat landasan teori yang mendukung penelitian, meliputi konsep smart gallery, metadata gambar, *cross-modal retrieval*, *embedding* teks dan gambar, Contrastive Language–Image Pre-trained (CLIP), metode CRISP-DM, Software Development Life Cycle (SDLC), serta kajian penelitian terdahulu yang relevan sebagai dasar ilmiah penelitian.

BAB III PEMBAHASAN

Bab ini membahas metodologi dan perancangan penelitian, meliputi rancangan penelitian, tahapan penelitian, sumber data, teknik pengumpulan data, teknik analisis data, perancangan model, evaluasi model, implementasi sistem smart gallery berbasis web, jadwal pelaksanaan penelitian, serta rincian biaya yang diperlukan dalam penelitian.

BAB IV HASIL DAN ANALISIS

Bab ini berisi hasil implementasi dan pengujian sistem yang telah dikembangkan. Pembahasan meliputi hasil fine-tuning model, hasil pengujian pencarian gambar berbasis teks, evaluasi performa sistem menggunakan metrik recall dan serta analisis terhadap efektivitas sistem dalam mendukung pencarian gambar berbasis *cross-modal retrieval*.

BAB V PENUTUP

Bab ini memuat kesimpulan yang diperoleh berdasarkan hasil penelitian serta saran yang dapat digunakan sebagai acuan untuk pengembangan sistem dan penelitian selanjutnya.

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

BAB 5

PENUTUP

5.1 Kesimpulan

Penelitian ini telah berhasil membandingkan dan mengimplementasikan pasangan model CLIP terbaik pada sistem pencarian gambar dengan teks bahasa Indonesia. Penelitian ini membuktikan adanya perbedaan performa yang signifikan di antara ketiga konfigurasi *text encoder* untuk tugas pencarian gambar dengan kueri berbahasa Indonesia dengan menggunakan dataset Flickr8K dan Flickr30K. Penelitian ini menghasilkan ranking yang konsisten yaitu dengan CLIP standar memberikan performa yang terbaik dibandingkan model CLIP dengan XLM-RoBERTa dan model CLIP dengan IndoBERT. Fitur yang berhasil dikembangkan model CLIP standar mencakup:

- 1) Pencarian gambar berbasis Teks: Sistem dapat menemukan gambar yang relevan secara semantik hanya melalui deskripsi teks berbahasa Indonesia sehari-hari.
- 2) Pencarian gambar berbasis Gambar: Sistem dapat menemukan gambar-gambar serupa di dalam galeri berdasarkan kemiripan fitur visual dari gambar referensi yang diunggah.

Hasil eksperimen secara keseluruhan mengindikasikan bahwa keberhasilan suatu model dalam *contrastive cross-modal retrieval* ditentukan oleh tiga faktor utama:

- 1) Keselarasan *embedding space*: model yang ruang representasinya sudah terselaraskan secara lintas modalitas sejak tahap *pre-training* terbukti lebih unggul dibandingkan model yang menggabungkan dua *embedding space* berbeda melalui *projection layer*.
- 2) Kapasitas transfer lintas bahasa: model dengan kemampuan generalisasi lintas bahasa yang lebih kuat menunjukkan adaptasi yang lebih efektif terhadap kueri Bahasa Indonesia dibandingkan model monolingual. Hal ini mengindikasikan bahwa representasi multilingual lebih kompatibel dengan *visual embedding space* yang bersifat *language-agnostic*.



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

- 3) Skala dan kesesuaian data pelatihan: berkat gap performa antara CLIP standar dan model berbasis BERT yang menyempit secara konsisten dari Flickr8K ke Flickr30K mengindikasikan bahwa ketersediaan data yang lebih besar berpotensi menggeser keunggulan dari faktor arsitektur menuju faktor adaptasi linguistik.

Ketiga faktor ini secara kolektif menjelaskan mengapa keselarasan *embedding space* lebih dominan dibandingkan kompetensi linguistik spesifik bahasa dalam kondisi data yang terbatas.

Penelitian ini juga berhasil mengimplementasikan model *text encoder* terbaik, yaitu CLIP standar, ke dalam sistem Smart Gallery untuk mendukung pencarian gambar multimodal berbahasa Indonesia. Implementasi ini dikategorikan sebagai berhasil karena dapat diterima dan dioperasikan dengan baik oleh pengguna serta tergolong kategori *Excellent* sesuai dengan hasil pengujian SUS terhadap 11 responden.

5.2 Saran

Berdasarkan hasil penelitian ini, terdapat peluang pengembangan lebih lanjut yang dapat dilakukan di masa depan sebagai berikut:

- 1) Mengeksplorasi pendekatan *Knowledge Distillation* sebagai alternatif *contrastive learning* untuk menjembatani kesenjangan *embedding space* lintas arsitektur, serta melakukan perbandingan langsung dengan pendekatan M-CLIP dan CLIP-Adapter untuk memperoleh gambaran komparatif yang lebih komprehensif.
- 2) Pengembangan selanjutnya disarankan untuk melakukan eksplorasi *hyperparameter tuning* yang lebih sistematis, seperti variasi *learning rate*, *batch size*, dan *temperature* pada InfoNCE loss, guna mengidentifikasi konfigurasi optimal yang berpotensi meningkatkan performa masing-masing model secara lebih menyeluruh.
- 3) Mengembangkan atau memanfaatkan dataset yang lebih merepresentasikan konten visual spesifik Indonesia, guna meningkatkan relevansi sistem dalam skenario penggunaan yang lebih kontekstual.



DAFTAR PUSTAKA

- dos Santos, G. O., Moreira, D. A. B., Ferreira, A. I., Silva, J., Pereira, L., Bueno, P., Sousa, T., Maia, H., Da Silva, N., Colombini, E., Pedrini, H., & Avila, S. (2023). CAPIVARA: Cost-efficient approach for improving multilingual CLIP performance on low-resource languages. *Proceedings of the 3rd Workshop on Multi-lingual Representation Learning (MRL)*, 184–207. <https://doi.org/10.18653/v1/2023.mrl-1.15>
- Chen, G., Hou, L., Chen, Y., Dai, W., Shang, L., Jiang, X., Liu, Q., Pan, J., & Wang, W. (2023). mCLIP: Multilingual CLIP via cross-lingual transfer. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 13028–13043. <https://doi.org/10.18653/v1/2023.acl-long.728>
- Goring, S. (2025). CLIPSE – A Minimalistic CLIP-Based Image Search Engine for Research. 1–5. <https://doi.org/https://doi.org/10.48550/arXiv.2504.17643>
- Aske, K. (2023). (Mis) Matching Metadata : Improving Accessibility in Digital Visual Archives through the EyCon Project (Mis) Matching Metadata : Improving Accessibility in Digital Visual Archives through the EyCon Project. 16(4). <https://doi.org/10.1145/3594726>
- Mohammadi, M. (2023). Image-Text Connection : Exploring the Expansion of the Diversity within Joint Feature Space Similarity Scores. *IEEE Access*, PP, 1. <https://doi.org/10.1109/ACCESS.2023.3327339>
- Gastaldi, J. L., Terilla, J., Malagutti, L., DuSell, B., Vieira, T., & Cotterell, R. (2025). *The foundations of tokenization: Statistical and computational concerns*. *Proceedings of the International Conference on Learning Representations (ICLR 2025)*. <https://arxiv.org/abs/2407.11606>
- Yi, W., Cai, X., Ma, H., Fu, Z., & Zhan, Y. (2025). *A library-oriented large language model approach to cross-lingual and cross-modal document retrieval*. *Electronics*, 14(15), 3145. <https://doi.org/10.3390/electronics14153145>

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

Gomes, G., Zerva, C., & Martins, B. (2024). Evaluation of Multilingual Image Captioning: How far can we get with CLIP models ? <https://doi.org/https://doi.org/10.48550/arXiv.2502.06600>

Han, Z., & Azman, A. B. I. N. (2024). Cross-Modal Retrieval : A Review of Methodologies , Datasets , and Future Perspectives. 12(August). <https://doi.org/https://doi.org/10.1109/ACCESS.2024.3444817>

Faisal, M. R., Nugrahad, D. T., & Budiman, I. (2024). 3D word *embedding* vector feature extraction and hybrid CNN- LSTM for natural disaster reports identification. 22(5), 1196–1208. <https://doi.org/10.12928/TELKOMNIKA.v22i5.26091>

Wang, T., Li, F., Zhu, L., Li, J., Zhang, Z., & Shen, H. T. (2024). Cross-Modal Retrieval : A Systematic Review of Methods and Future Directions. 1–35. <https://doi.org/https://doi.org/10.1109/JPROC.2024.3525147>

Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2022). *Efficient transformers: A survey*. *ACM Computing Surveys*, 55(6), Article 109, 1–38. <https://doi.org/10.1145/3530811>

Radford, A., Wook, J., Chris, K., Aditya, H., Gabriel, R., Sandhini, G., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. <https://doi.org/10.48550/arXiv.2103.00020>

Hwang, T., Seo, H., Jung, J., & Jung, S. (2025). *Exploring selective layer freezing strategies in transformer fine-tuning: NLI classifiers with sub-3B parameter models*. *Applied Sciences*, 15(19), 10434. <https://doi.org/10.3390/app151910434>

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale. ICLR 2021 - 9th International Conference on Learning Representations. <https://doi.org/https://doi.org/10.48550/arXiv.2010.11929>



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

Huang, Z., Dong, Q., Gong, S., & Zhu, X. (2023). *Model-aware contrastive learning: Towards escaping the dilemmas of InfoNCE*. Proceedings of the 40th International Conference on Machine Learning (ICML).

<https://arxiv.org/abs/2306.13298>

Abasova, J., Tanuska, P., & Rydzi, S. (2021). Big data—knowledge discovery in production industry data storages—implementation of best practices. *Applied Sciences (Switzerland)*, 11(16). <https://doi.org/10.3390/app11167648>

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). *Unsupervised cross-lingual representation learning at scale*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 8440–8451. <https://doi.org/10.18653/v1/2020.acl-main.747>

Mohammadi, M., Eftekhari, M., & Hassani, A. (2023). Image-text connection: Exploring the expansion of the diversity within joint feature space similarity scores. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2023.3327339>

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP*. arXiv. <https://arxiv.org/abs/2011.00677>

Liang, W., Zhang, Y., Kwon, Y., Yeung, S., & Zou, J. (2022). Mind the gap: Understanding the modality gap in multi-modal contrastive representation learning. *Advances in Neural Information Processing Systems*, 35, 17612–17625. <https://doi.org/10.48550/arXiv.2203.02053>

Wortsman, M., Ilharco, G., Kim, J. W., Li, M., Kornblith, S., Roelofs, R., Gontijo-Lopes, R., Hajishirzi, H., Farhadi, A., Namkoong, H., & Schmidt, L. (2022). *Robust fine-tuning of zero-shot models*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7959–7971. DOI: <https://doi.org/10.1109/CVPR52688.2022.00780>



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

Pambudi, A., & Abidin, Z. (2023). PENERAPAN CRISP - DM MENGGUNAKAN MLR K - FOLD PADA DATA SAHAM PT . TELKOM INDONESIA (PERSERO) TBK (TLKM) (STUDI KASUS : BURSA EFEK. 4(1), 1–14. <https://doi.org/https://doi.org/10.33365/jdmsi.v4i1.2462>

Plotnikova, V., Dumas, M., & Milani, F. P. (2022). Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data and Knowledge Engineering*, 139(October 2021), 1–17. <https://doi.org/10.1016/j.datak.2022.102013>

Martinez-Plumed, F., Contreras-ochando, L., Kull, M., & Lachiche, N. (2021). CRISP-DM Twenty Years Later : From Data Mining Processes to Data Science Trajectories. 4347(c). <https://doi.org/10.1109/TKDE.2019.2962680>

Kong, J. T. H., Juwono, F. H., Ngu, I. Y., Nugraha, I. G. D., Maraden, Y., & Wong, W. K. (2023). A Mixed Malay – English Language COVID-19 Twitter Dataset : A Sentiment Analysis. <https://doi.org/https://doi.org/10.3390/bdcc7020061>

Maina, N. K., Muketha, G. M., & Wambugu, G. M. (2023). *A new complexity metric for UML sequence diagrams. International Journal of Software Engineering & Applications*, 14(1), 9–22. <https://doi.org/10.5121/ijsea.2023.14102>

Lapo, J. R., & Cumbicus-Pineda, O. M. (2024). Detection of recyclable solid waste using convolutional neural networks and PyTorch. *IEEE Latin America Transactions*, 22(6), 475–483. <https://doi.org/10.1109/TLA.2024.10534304>

Ramzi, E., & Audebert, N. (2023). Optimization of Rank Losses for Image Retrieval. 1–20. <https://doi.org/https://doi.org/10.48550/arXiv.2309.08250>

Yas, Q. M., Alazzawi, A., Yas, Q. M., & Rahmatullah, B. (2023). A Comprehensive Review of Software Development Life Cycle methodologies : Pros , Cons , and Future Directions A Comprehensive Review of Software Development Life Cycle methodologies : Pros , Cons , and Future Directions. 4(4). <https://doi.org/https://doi.org/10.52866/ijcsm.2023.04.04.014>



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

Ghumatkar, R. S., & Date, A. (2023). Software Development Life Cycle (SDLC). November. <https://doi.org/https://doi.org/10.22214/ijraset.2023.56554>

Huang, X., Fayoumi, A., Winter, E., & Najdawi, A. (2025). Design Requirements of Breast Cancer Symptom- Management Apps. 1–24. <https://doi.org/https://doi.org/10.3390/informatics12030072>

Zhou, D.-W., Cai, Z.-W., Ye, H.-J., Zhan, D.-C., & Liu, Z. (2024). *Revisiting class-incremental learning with pre-trained models: Generalizability and adaptivity are all you need. International Journal of Computer Vision.* <https://doi.org/10.1007/s11263-024-02220-5>

Nguyen, T.-T., Huynh, V.-T., Nguyen, Q.-T., Nguyen, H.-P., Bao, L. L., Minh, T. H., Anh, M. N., Tien, T. N., Thuan, P. N., Phong, H. N., Thai, B. H., Nguyen, V.-T., Nguyen, D.-V., Pham, P.-H., Le-Hoang, M.-H., Le, N.-K., Nguyen, M.-C., Ho, M.-Q., Tran, N.-L., ... Tran, M.-T. (2025). *SHREC 2025: Retrieval of optimal objects for multi-modal enhanced language and spatial assistance (ROOMELSA).* *Computers & Graphics*, 132, 104400. <https://doi.org/10.1016/j.cag.2025.104400>

Fumero, M., Pegoraro, M., Maiorca, V., Locatello, F., & Rodolà, E. (2024). *Latent functional maps: A spectral framework for representation alignment.* In *Advances in Neural Information Processing Systems* (Vol. 37). <https://doi.org/10.52202/079017-2115>

Rusialdi, F. M. (n.d.). Rancang Bangun Aplikasi Jasa Titip Lokal Berbasis Web Dengan Integrasi Pembayaran Midtrans. Multinetics.

Zhai, X., Wang, X., Mustafa, B., Steiner, A., Keysers, D., Kolesnikov, A., & Beyer, L. (2022). *LiT: Zero-shot transfer with locked-image text tuning.* In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 18123–18133). <https://doi.org/10.1109/CVPR52688.2022.01762>

Unas, B. M., & Jurgelaitis, M. (2024). Transforming Sketches of UML Use Case Diagrams to Models. 12(October). <https://doi.org/10.1109/ACCESS.2024.3514455>



© 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

Bianchi, F., Attanasio, G., Pisoni, R., Terragni, S., Sarti, G., & Lakshmi, S. (2021). *Contrastive language-image pre-training for the Italian language*. arXiv.

<https://arxiv.org/abs/2108.08688>

Alyami, A., Pileggi, S. F., Sohaib, O., & Hawryszkiewicz, I. (2023). Seamless transformation from use case to sequence diagrams. 1–26.

<https://doi.org/10.7717/peerj-cs.1444>

Gong, Y., Cosma, G., & Finke, A. (2023). Neural-Based Cross-Modal Search And Retrieval Of Artwork.

<https://doi.org/https://doi.org/10.1109/SSCI52147.2023.10371948>



DAFTAR RIWAYAT HIDUP



Muhammad Nur Irfan

Lahir di Jakarta pada tanggal 26 Juni 2004, sebagai anak pertama dari dua bersaudara. Lulus dari SD Angkasa 7 pada tahun 2016, kemudian melanjutkan pendidikan di SMP Negeri 80 Jakarta dan lulus pada tahun 2019, serta menempuh pendidikan di SMA Negeri 67 Jakarta pada tahun 2022. Saat ini menempuh Pendidikan Sarjana Terapan pada Program Studi Teknik Informatika di Politeknik Negeri Jakarta pada tahun 2022.

© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

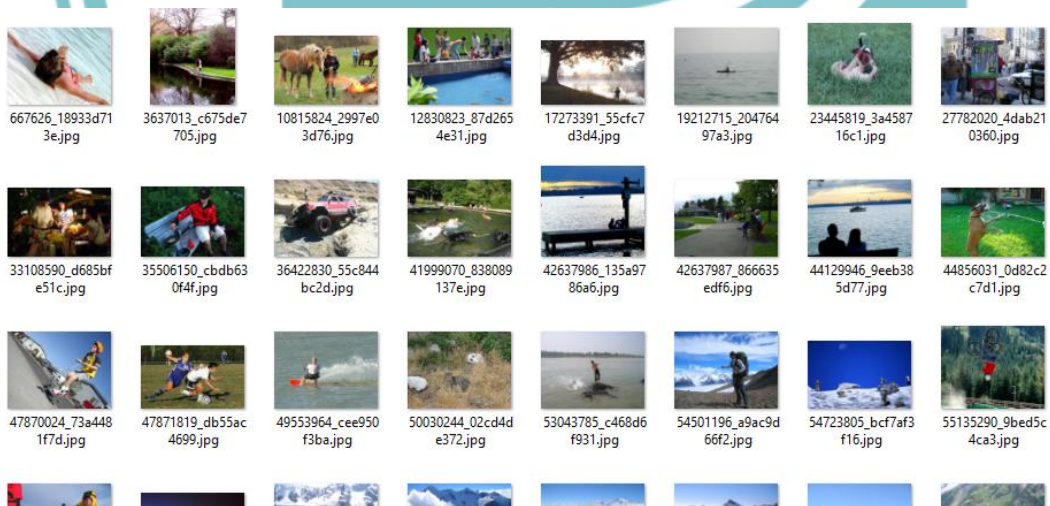
Hak Cipta :

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 mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**POLITEKNIK
NEGERI
JAKARTA**

Lampiran 1 Contoh Dataset Flickr8K Indonesia

[illegible]

[illegible]



Hak Cipta :

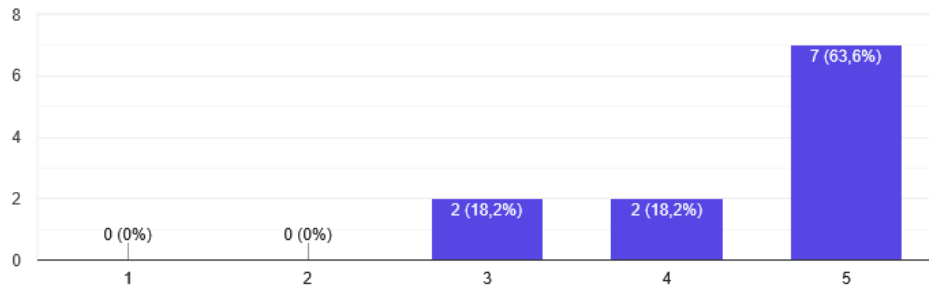
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

Lampiran 3 Pertanyaan Kuesioner SUS

Saya berpikir akan menggunakan web ini lagi

[Salin diagram](#)

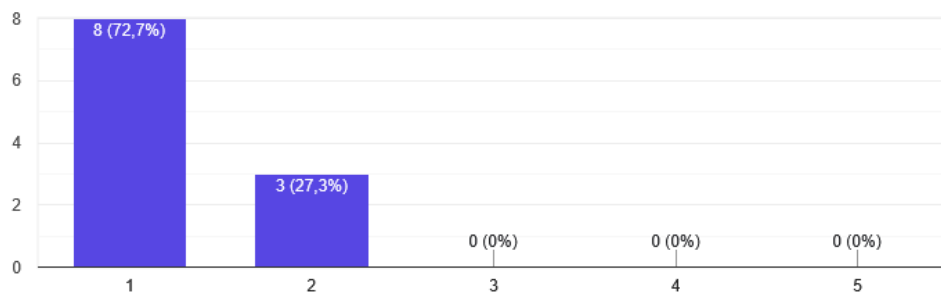
11 jawaban



Saya merasa website ini rumit untuk digunakan

[Salin diagram](#)

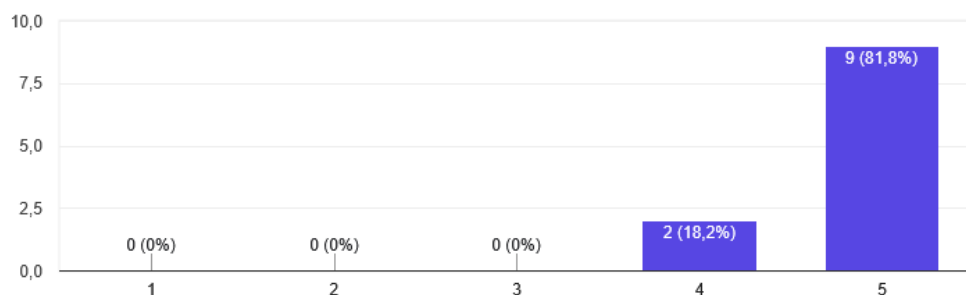
11 jawaban



Saya merasa website ini mudah untuk digunakan

[Salin diagram](#)

11 jawaban



(Lanjutan)



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

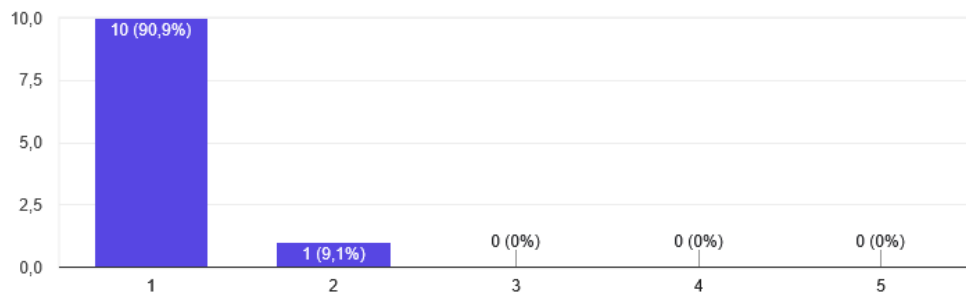
Hak Cipta :

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

Saya membutuhkan orang lain atau teknisi untuk menggunakan website ini

[Salin diagram](#)

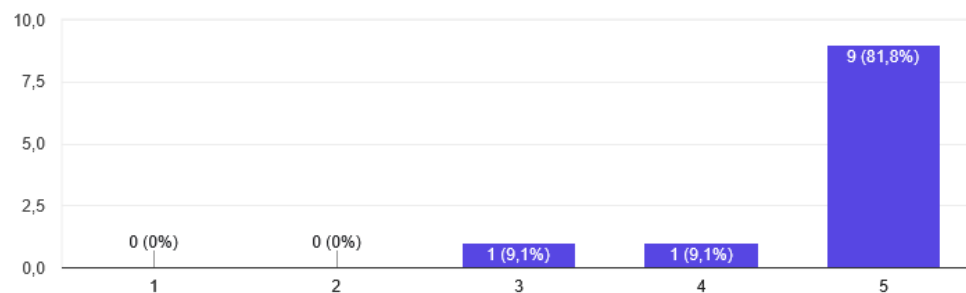
11 jawaban



Saya merasa fitur-fitur yang tersedia berjalan dengan semestinya

[Salin diagram](#)

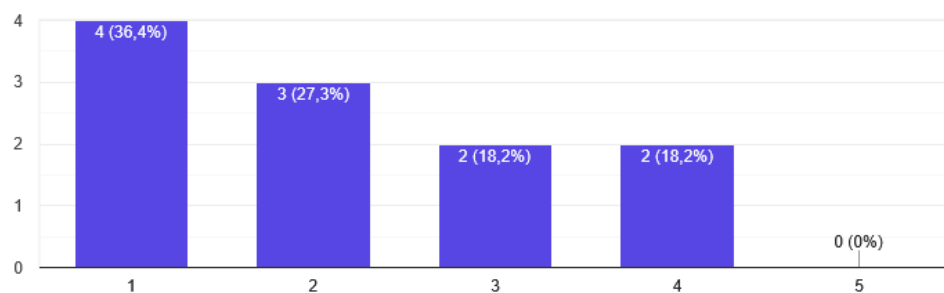
11 jawaban



Saya merasa ada banyak hal yang tidak konsisten (tidak teratur) pada website ini

[Salin diagram](#)

11 jawaban



(Lanjutan)



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

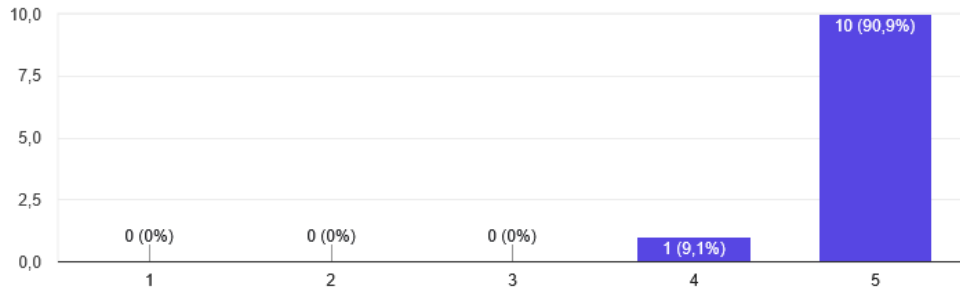
Hak Cipta :

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 mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Saya rasa orang lain dapat memahami cara menggunakan website ini dengan cepat

[Salin diagram](#)

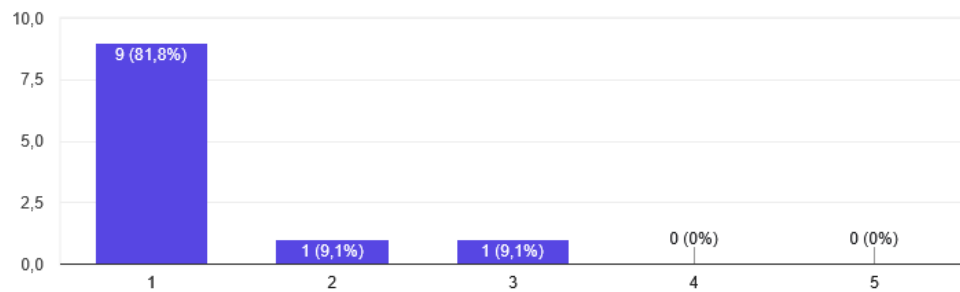
11 jawaban



Saya merasa website pencarian gambar ini membingungkan

[Salin diagram](#)

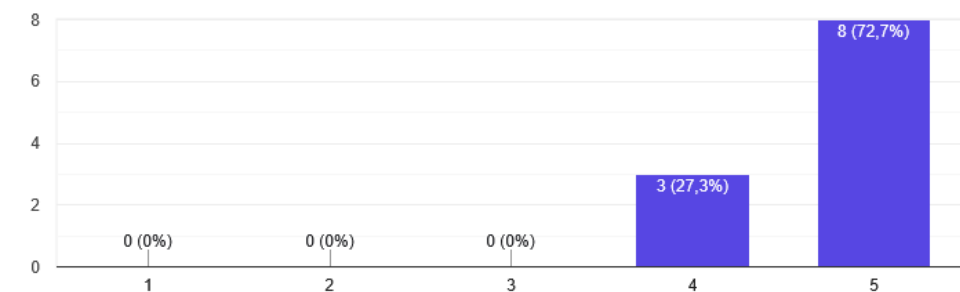
11 jawaban



Saya merasa tidak memiliki hambatan ketika menggunakan website ini

[Salin diagram](#)

11 jawaban

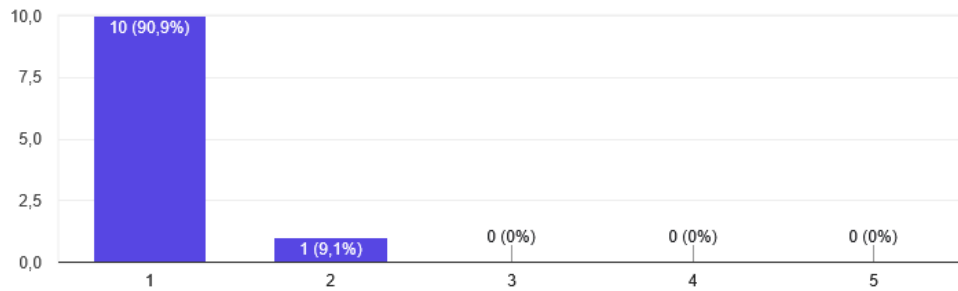


(Lanjutan)

 Salin diagram

Saya perlu membiasakan diri terlebih dahulu untuk menggunakan website ini

11 jawaban



© 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 menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta