

IMPLEMENTASI MULTI-LABEL TEXT CLASSIFICATION MENGUNAKAN INDOBERT UNTUK KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS BERBAHASA INDONESIA

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh :

Bariq Khairullah

NIM : 09021282126094

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2025

HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI MULTI-LABEL TEXT CLASSIFICATION
MENGUNAKAN INDOBERT UNTUK KLASIFIKASI GENRE FILM
BERDASARKAN SINOPSIS BERBAHASA INDONESIA

Sebagai salah satu syarat untuk penyelesaian studi di
Program Studi S1 Teknik Informatika

Oleh:
BARIQ KHAIRULLAH
09021282126094

Pembimbing 1 : Dr. Abdiansah, S.Kom., M.CS.
NIP. 198410012009121005

Mengetahui
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D
198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari tanggal telah dilaksanakan ujian komprhensif
skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas
Sriwijaya.

Nama : Bariq Khairullah

NIM : 09021282126094

Judul : Implementasi Multi-Label Text Classification
Menggunakan Indobert Untuk Klasifikasi Genre Film
Berdasarkan Sinopsis Berbahasa Indonesia

Dan dinyatakan

1. Ketua Penguji

Yunita, M.Cs.
NIP. 198306062015042002

2. Penguji

Rifkie Primartha, M.T.
NIP. 197706012009121004

3. Pembimbing

Dr. Abdiansah, S.Kom., M.CS.
NIP. 198410012009121005



Mengetahui

Ketua Jurusan Teknik Informatika

Hadipurnawan Satna, S.Kom., M.Sc., Ph.D.

NIP. 198004182020121001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Bariq Khairullah
NIM : 09021282126094
Judul : Implementasi Multi-Label Text Classification
Menggunakan Indobert Untuk Klasifikasi Genre Film
Berdasarkan Sinopsis Berbahasa Indonesia

Hasil Pengecekan Software Turnitin: 3%

Menyertakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak manapun.



Palembang, 28 Februari 2025



Bariq Khairullah

NIM. 090212826094

MOTO DAN PERSEMBAHAN

"Berubah, beradaptasi, berkembang—karena dunia tidak menunggu siapa pun."

(Kaizen Ardent)

Kupersembahkan karya tulis ini
kepada :

- Diriku
- Orang Tua dan Saudaraku
- Teman Seperjuanganku

ABSTRACT

This research is motivated by the rapid growth of the Indonesian film industry and the increasing need for an accurate film genre classification system to help viewers understand film content more precisely. Synopsis is often the main reference in identifying film genres, but the complexity and diversity of languages make the classification process a challenge. To answer this challenge, this study carries IndoBERT as the main method in classifying film genres in a multi-label manner based on Indonesian synopses. IndoBERT was chosen because of its ability to understand the context and structure of the Indonesian language in depth. The dataset used consists of 1,738 films with five main genres: Drama, Comedy, Horror, Action, and Romance. As a complement, this study also applies two non-core optimization techniques, namely Dynamic Thresholding and Per-class Performance Tracking, to overcome data imbalances between genres and improve prediction accuracy. As a result, the IndoBERT model managed to achieve an accuracy of 79.23% and a Macro F1-score of 57.15%, an increase of 6.21% from the initial baseline of 73.03%. The horror genre recorded the best performance among the other genres with an accuracy of 88.89% and an F1-score of 77.17%, supported by distinctive and consistent linguistic features. The model was integrated into a Streamlit-based web application with an average response time of under 0.25 seconds, proving that IndoBERT is effective and efficient for movie genre classification in real-world scenarios.

Keywords: IndoBERT, Multi-label Classification, Film Genre, Natural Language Processing, and Film Synopsis.

ABSTRAK

Penelitian ini dilatarbelakangi oleh pesatnya pertumbuhan industri perfilman Indonesia dan meningkatnya kebutuhan akan sistem klasifikasi genre film yang akurat untuk membantu penonton memahami konten film dengan lebih tepat. Sinopsis sering kali menjadi referensi utama dalam mengidentifikasi genre film, namun kompleksitas dan keragaman bahasa membuat proses klasifikasi menjadi tantangan tersendiri. Untuk menjawab tantangan tersebut, penelitian ini mengusung IndoBERT sebagai metode utama dalam mengklasifikasikan genre film secara multi-label berdasarkan sinopsis berbahasa Indonesia. IndoBERT dipilih karena kemampuannya dalam memahami konteks dan struktur bahasa Indonesia secara mendalam. Dataset yang digunakan terdiri dari 1.738 film dengan lima genre utama: Drama, Komedi, Horor, Laga, dan Romantis. Sebagai pelengkap, penelitian ini juga menerapkan dua teknik optimasi non-inti, yaitu Dynamic Thresholding dan Per-class Performance Tracking, guna mengatasi ketidakseimbangan data antar genre dan meningkatkan ketepatan prediksi. Hasilnya, model IndoBERT berhasil mencapai akurasi 79,23% dan Macro F1-score 57,15%, meningkat 6,21% dari baseline awal 73,03%. Genre horor mencatatkan performa terbaik dari genre yang lain dengan akurasi 88,89% dan F1-score 77,17%, didukung oleh ciri linguistik yang khas dan konsisten. Model ini diintegrasikan ke dalam aplikasi web berbasis Streamlit dengan rata-rata waktu respons di bawah 0,25 detik, membuktikan bahwa IndoBERT efektif dan efisien untuk klasifikasi genre film dalam skenario dunia nyata.

Keywords: IndoBERT, Multi-label Classification, Genre Film, Natural Language Processing, dan Sinopsis Film.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT atas segala limpahan berkah, rahmat, dan hidayah-Nya sehingga penulis dapat melaksanakan dan menyelesaikan Tugas Akhir ini yang berjudul *"Implementasi Multi-Label Text Classification Menggunakan IndoBERT untuk Klasifikasi Genre Film Berdasarkan Sinopsis Berbahasa Indonesia."* Tugas Akhir ini disusun sebagai salah satu syarat untuk penyelesaian studi di Program Studi Sistem Informasi Reguler Universitas Sriwijaya.

Dalam proses penyelesaian Tugas Akhir ini, penulis banyak dibantu oleh berbagai pihak. Bantuan tersebut berupa bimbingan, pengarahan, nasihat, dan pemikiran dalam menyelesaikan Tugas Akhir ini. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Erwin, S.Si., M.Si. sebagai Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Bapak Hadipurniawan Satria, M.Sc., Ph.D. sebagai Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Dr. Abdiansah, S.Kom., M.CS. selaku Dosen Pembimbing Tugas Akhir yang telah bersedia meluangkan waktunya untuk memberikan bimbingan, saran, dan kritik yang membangun dalam penyusunan laporan tugas akhir ini.

4. Dosen-dosen Universitas Sriwijaya, terutama dosen Teknik Informatika.
5. Kedua orang tua, dan adik yang selalu memberikan doa, semangat, serta bantuan baik materi maupun non-materi sehingga laporan ini dapat terselesaikan.
6. Seluruh teman-teman Teknik Informatika Bilingual dan Reguler angkatan 2021 serta seluruh pihak yang tidak dapat saya sebutkan satu per satu.

Tugas akhir ini merupakan hasil kerja yang seoptimal mungkin. Namun, penulis menyadari bahwa dalam penulisan Tugas Akhir ini masih terdapat banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik, saran, serta masukan yang sifatnya membangun dari semua pihak demi kesempurnaan di masa yang akan datang.

Penulis berharap semoga Tugas Akhir ini bermanfaat bagi penulis sendiri dan semua pihak.

Indralaya, 12 Februari 2025

Penyusun

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	Error! Bookmark not defined.
TANDA LULUS UJIAN KOMPREHENSIF	Error! Bookmark not defined.
HALAMAN PERNYATAAN	Error! Bookmark not defined.
MOTO DAN PERSEMBAHAN.....	iv
ABSTRACT.....	v
ABSTRAK.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah.....	I-3
1.4 Tujuan Penelitian.....	I-3
1.5 Manfaat Penelitian	I-3
1.6 Batasan Masalah	I-4
1.7 Sistematika Penulisan.....	I-4
1.8 Kesimpulan.....	I-5
BAB II KAJIAN LITERATUR.....	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Sinopsis Film.....	II-1
2.2.2 Genre Film	II-1
2.2.3 Multi-label Classification	II-2
2.2.4 Natural Language Processing	II-3
2.2.5 Model IndoBERT.....	II-7
2.2.6 Metrics Evaluasi	II-10
2.2.7 Streamlit.....	II-15
2.3 Penelitian Terdahulu	II-15

2.4 Kesimpulan.....	II-20
BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data.....	III-1
3.2.1 FilmIndonesia.....	III-1
3.2.2 Themoviedb	III-3
3.2.3 Kaggle Dataset IMDb Indonesian Movies	III-4
3.2.4 Kaggle Dataset IMDb Synopsis Indonesian Movies	III-5
3.2.5 Intergrasi Data.....	III-6
3.3 Tahapan Penelitian.....	III-8
3.3.1 Kerangka Kerja Penelitian.....	III-9
3.3.2 Kriteria Pengujian.....	III-13
3.3.3 Format Data Pengujian.....	III-15
3.3.4 Alat Bantu Penelitian.....	III-19
3.3.5 Alat Ukur Pengujian.....	III-20
3.3.6 Analisis Hasil Pengujian dan Menarik Kesimpulan.....	III-21
3.4 Kesimpulan.....	III-24
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 Fase Analisis	IV-2
4.2.1 Analisis Kebutuhan.....	IV-2
4.2.2 Analisis Sistem Perangkat Lunak	IV-7
4.3 Fase Pengumpulan Dataset.....	IV-9
4.4 Fase Perancangan Sistem	IV-13
4.4.1 Perancangan Perangkat Lunak	IV-14
4.4.2 Perancangan Antarmuka	IV-35
4.4.3 Perancangan Model	IV-41
4.5 Fase Implementasi	IV-42
4.5.1 Implementasi Antarmuka.....	IV-43
4.5.2 Implementasi Pelatihan Model	IV-52
4.5.3 Implementasi Integrasi Antarmuka dan Model.....	IV-54
4.6 Kesimpulan.....	IV-56
BAB V HASIL DAN PEMBAHASAN.....	V-1

5.1 Pendahuluan	V-1
5.2 Hasil Pengujian Model.....	V-1
5.2.1 Distribusi Dataset	V-3
5.2.2 Konfigurasi Pengujian.....	V-3
5.2.3 Hasil Evaluasi Overall.....	V-5
5.2.4 -Uji Performa Real-World Prediction Pada Data Eksternal.....	V-9
5.2.5 Analisis Performa Optimasi	V-14
5.2.6 Analisis Performa Overall	V-15
5.2.7 Analisis Performa Per-Genre	V-17
5.2.8 Analisis Confusion Metrics Per-Genre.....	V-41
5.3 Hasil Pengujian Aplikasi.....	V-59
5.3.1 Hasil Uji Fungsionalitas	V-60
5.3.2 Hasil Uji Performa Berdasarkan Panjang Sinopsis	V-62
5.4 Kesimpulan.....	V-63
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Kesimpulan.....	VI-1
6.2 Saran	VI-2
DAFTAR PUSTAKA	xvii
LAMPIRAN.....	xxi

DAFTAR TABEL

Tabel II- 1 Contoh Multi-label Classification	II-3
Tabel II- 2 Contoh Text Preprocessing Pada NLP	II-4
Tabel II- 3 Contoh Tokenisasi Pada NLP	II-5
Tabel II- 4 Contoh Stemming Pada NLP	II-5
Tabel II- 5 Contoh Lemmatization Pada NLP	II-6
Tabel II- 6 Contoh Text Representation Pada NLP	II-6
Tabel II- 7 Perbandingan Arsitektur IndoBERT	II-10
Tabel II- 8 Penelitian Terdahulu	II-16
Tabel III- 1 Rancangan Tabel Hasil Evaluasi Overall	III-21
Tabel III- 2 Rancangan Tabel Hasil Evaluasi Per-Genre	III-23
Tabel IV- 1 Sample Datasets	IV-10
Tabel IV- 2 Sampel Data Sebelum Preprocessing Teks	IV-23
Tabel IV- 3 Sampel Data Setelah Data Cleaning	IV-24
Tabel IV- 4 Sampel Data Setelah Case Folding	IV-26
Tabel IV- 5 Sampel Datasets Setelah Stopwords Removal	IV-29
Tabel IV- 6 Tabel Sampel Dataset Setelah Tokenisasi	IV-30
Tabel V- 1 Konfigurasi Default	V-4
Tabel V- 2 Konfigurasi 1 Hyperparameter Tuning	V-4
Tabel V- 3 Konfigurasi 2 Hyperparameter Tuning	V-5
Tabel V- 4 Hasil Evaluasi Overall Pengujian	V-6
Tabel V- 5 Performa Per-Genre Pengujian 1	V-17
Tabel V- 6 Performa Per-Genre Pengujian 2	V-19

Tabel V- 7 Performa Per-Genre Pengujian 3	V-22
Tabel V- 8 Performa Per-Genre Pengujian 4	V-24
Tabel V- 9 Performa Per-Genre Pengujian 5	V-27
Tabel V- 10 Performa Per-Genre Pengujian 6	V-30
Tabel V- 11 Performa Per-Genre Pengujian 7	V-32
Tabel V- 12 Performa Per-Genre Pengujian 8	V-35
Tabel V- 13 Hasil Uji Fungsionalitas.....	V-60
Tabel V- 14 Hasil Uji Performa Berdasarkan Panjang Sinopsis.....	V-62

DAFTAR GAMBAR

Gambar II- 1 Contoh Arsitektur IndoBERT	II-8
Gambar III- 1 Grafik Distribusi Datasets FilmIndonesia	III-2
Gambar III- 2 Grafik Distribusi Datasets Themoviedb	III-4
Gambar III- 3 Grafik Distribusi Datasets Kaggle IMDb Indonesian Movies..	III-5
Gambar III- 4 Grafik Distribusi Datasets IMDB Synopsis Indonesian Movies	III-6
Gambar III- 5 Grafik Distribusi Datasets Setelah di Intergrasi	III-8
Gambar III- 6 Tahapan Penelitian.....	III-9
Gambar III- 7 Kerangka Kerja Penelitian	III-10
Gambar III- 8 Header Performa Overall	III-16
Gambar III- 9 Tabel Performa Per-Genre	III-17
Gambar III- 10 Grid Layout Confusion Metrics.....	III-18
Gambar III- 11 Template Confusion Metrics	III-18
Gambar IV- 1 Diagram Alur Sistem.....	IV-7
Gambar IV- 2 Use Case Diagram Sistem Klasifikasi.....	IV-15
Gambar IV- 3 Flowchart Diagram Alur Prediksi Sinopsis	IV-20
Gambar IV- 4 Sequence Diagram Sistem Aplikasi	IV-33
Gambar IV- 5 Rancangan Antarmuka Tab Model Overview	IV-36
Gambar IV- 6 Rancangan Antarmuka Tab Training Metrics.....	IV-37
Gambar IV- 7 Rancangan Antarmuka Tab Training Logs	IV-38
Gambar IV- 8 Rancangan Antarmuka Tab Evaluation Result.....	IV-38

Gambar IV- 9	Rancangan Antarmuka Tab Make Prediction	IV-39
Gambar IV- 10	Rancangan Antarmuka Tab Make Prediction dengan Hasil ...	IV-40
Gambar IV- 11	Rancangan Antarmuka Tab Datasets Information	IV-41
Gambar IV- 12	Perancangan Model	IV-41
Gambar IV- 13	Antarmuka Tab Model Overview 1	IV-43
Gambar IV- 14	Antarmuka Tab Model Overview 2	IV-43
Gambar IV- 15	Antarmuka Tab Training Metrics 1	IV-44
Gambar IV- 16	Antarmuka Tab Training Metrics 2	IV-45
Gambar IV- 17	Antarmuka Tab Training logs	IV-46
Gambar IV- 18	Antarmuka Tab Evaluation Result 1	IV-47
Gambar IV- 19	Antarmuka Tab Evaluation Result 2.....	IV-47
Gambar IV- 20	Antarmuka Tab Make Prediction tanpa Output	IV-48
Gambar IV- 21	Antarmuka Tab Make Prediction dengan Output	IV-49
Gambar IV- 22	Antarmuka Tab Dataset Information	IV-50
Gambar IV- 23	Antarmuka Tab Resource Information	IV-51
Gambar IV- 24	Antarmuka Tab System Info	IV-52
Gambar V- 1	Grafik Hasil Evaluasi Overall Pengujian.....	V-7
Gambar V- 2	Grafik Performa Per-Genre Pengujian 1	V-18
Gambar V- 3	Grafik Performa Per-Genre Pengujian 2.....	V-20
Gambar V- 4	Grafik Performa Per-Genre Pengujian 3	V-22
Gambar V- 5	Grafik Performa Per-Genre Pengujian 4.....	V-25
Gambar V- 6	Grafik Performa Per-Genre Pengujian 5	V-28

Gambar V- 7 Grafik Performa Per-Genre Pengujian 6	V-30
Gambar V- 8 Grafik Performa Per-Genre Pengujian 7	V-33
Gambar V- 9 Grafik Performa Per-Genre Pengujian 8	V-36
Gambar V- 10 Confusion Metrics Genre Drama Pengujian 1	V-42
Gambar V- 11 Confusion Metrics Genre Horor Pengujian 1	V-43
Gambar V- 12 Confusion Metrics Genre Drama Pengujian 2	V-44
Gambar V- 13 Confusion Metrics Genre Horor Pengujian 2	V-45
Gambar V- 14 Confusion Metrics Genre Drama Pengujian 3	V-46
Gambar V- 15 Confusion Metrics Genre Horor Pengujian 3	V-47
Gambar V- 16 Confusion Metrics Genre Drama Pengujian 4	V-48
Gambar V- 17 Confusion Metrics Genre Horor Pengujian 4	V-49
Gambar V- 18 Confusion Metrics Genre Drama Pengujian 5	V-50
Gambar V- 19 Confusion Metrics Genre Horor Pengujian 5	V-51
Gambar V- 20 Confusion Metrics Genre Drama Pengujian 6	V-52
Gambar V- 21 Confusion Metrics Genre Horor Pengujian 6	V-53
Gambar V- 22 Confusion Metrics Genre Drama Pengujian 7	V-54
Gambar V- 23 Confusion Metrics Genre Horor Pengujian 7	V-55
Gambar V- 24 Confusion Metrics Genre Drama Pengujian 8	V-56
Gambar V- 25 Confusion Metrics Genre Horor Pengujian 8	V-57

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini akan menguraikan elemen-elemen utama yang menjadi landasan penelitian ini. Pembahasan akan mencakup beberapa aspek kunci, di antaranya latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian. Uraian mengenai aspek-aspek fundamental tersebut akan berfungsi sebagai kerangka dasar yang mengarahkan keseluruhan proses penelitian ini.

1.2 Latar Belakang Masalah

Pada kurun waktu lima tahun terakhir, pertumbuhan Perfilman Nasional menunjukkan perkembangan yang menggembirakan. Tren positif ini terlihat dari beberapa indikator utama, antara lain jumlah produksi film dan jumlah penonton. Berdasarkan data Pusat Pengembangan Perfilman Indonesia, produksi film nasional pada tahun 2018 mencapai 132 judul dengan 51,2 juta penonton, sementara tahun 2019 mencatatkan 129 judul dengan jumlah penonton yang relatif stabil di angka 51,2 juta (Badan Perfilman Indonesia [BPI], 2025).

Seiring dengan pertumbuhan industri film nasional yang pesat, kebutuhan akan sistem klasifikasi film yang akurat menjadi semakin penting. Film memiliki beragam genre seperti romantis, horor, thriller, komedi, dan fantasi. Namun, masih banyak penikmat film yang mengalami kesulitan dalam membedakan genre-genre

tersebut, sehingga pesan yang ingin disampaikan melalui film tidak dapat tersampaikan secara maksimal kepada penontonnya (Akbar et al., 2025)

Berbagai penelitian telah dikembangkan untuk meningkatkan akurasi dalam klasifikasi genre film berdasarkan sinopsis. Studi oleh Madana (2021) membandingkan metode *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN) dalam mengklasifikasikan genre film berbahasa Indonesia. Hasil penelitian menunjukkan bahwa pendekatan berbasis deep learning dengan CNN dan Word2Vec memiliki performa lebih baik dibandingkan dengan SVM berbasis TF-IDF, dengan rata-rata F1-score masing-masing sebesar 64% dan 54%. Hal ini mengindikasikan bahwa metode deep learning dapat lebih efektif dalam menangkap pola bahasa alami dalam sinopsis film.

Selain itu, perkembangan teknologi *Natural Language Processing* (NLP) telah memberikan dampak signifikan terhadap sistem klasifikasi berbasis teks. Model bahasa pra-latih seperti *IndoBERT* yang diperkenalkan oleh (Koto et al., 2020a) telah terbukti unggul dalam berbagai tugas NLP di bahasa Indonesia. Wongso et al. (2024) kemudian mengembangkan *NusaBERT*, yang memperluas cakupan *IndoBERT* dengan memasukkan berbagai bahasa dan dialek regional di Indonesia. Penggunaan model berbasis transformer ini menunjukkan potensi besar dalam meningkatkan akurasi klasifikasi genre film berdasarkan sinopsis.

(disini penjelasan teknik optimasi)

Dengan mempertimbangkan hasil penelitian sebelumnya dan kemajuan teknologi NLP, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi genre film menggunakan pendekatan berbasis transformer, khususnya *IndoBERT*, guna meningkatkan akurasi klasifikasi berbasis teks sinopsis.

1.3 Rumusan Masalah

1. Bagaimana mengklasifikasikan Genre Film Multi-Label menggunakan IndoBERT?
2. Bagaimana kinerja model *IndoBERT* dalam mengklasifikasikan Genre Film Multi-Label?

1.4 Tujuan Penelitian

1. Menghasilkan aplikasi yang dapat mengklasifikasikan Genre Film Multi-Label menggunakan IndoBERT.
2. Mengetahui Kinerja Model IndoBERT dalam mengklasifikasikan Genre Film Multi-Label .

1.5 Manfaat Penelitian

1. Memberikan kontribusi dalam pengembangan teknik klasifikasi teks berbahasa Indonesia menggunakan metode berbasis *Natural Language Processing* (NLP).
2. Menyediakan analisis komparatif terhadap metode klasifikasi berbasis machine learning dan deep learning dalam klasifikasi genre film.
3. Mempermudah proses klasifikasi genre film secara otomatis berdasarkan sinopsis, sehingga mengurangi ketergantungan pada klasifikasi manual.

4. Menjadi dasar bagi pengembangan sistem rekomendasi film berbasis teks yang lebih kontekstual dan relevan dengan preferensi pengguna.

1.6 Batasan Masalah

1. Penelitian fokus pada klasifikasi teks berbahasa Indonesia.
2. Dataset terbatas pada sinopsis film.
3. Label genre yang digunakan terbatas pada lima genre utama, yaitu Drama, Komedi, Horor, Laga, dan Romantis.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

BAB I. PENDAHULUAN

Bab ini akan menguraikan elemen-elemen utama yang menjadi landasan penelitian ini. Pembahasan akan mencakup beberapa aspek kunci, di antaranya latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian. Uraian mengenai aspek-aspek fundamental tersebut akan berfungsi sebagai kerangka dasar yang mengarahkan keseluruhan proses penelitian ini.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dikaji dasar-dasar teori yang digunakan dalam penelitian ini serta juga membahas tentang penelitian yang relevan sebagai acuan dasar untuk penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dikaji mengenai perencanaan yang dilakukan untuk Bab ini membahas mengenai tahap-tahap yang akan diterapkan pada penelitian. Setiap rencana dari tahapan penelitian dideskripsikan secara rinci berdasarkan kerangka kerja. Dilanjutkan dengan perancangan manajemen proyek dalam pelaksanaan penelitian.

BAB IV. IMPLEMENTASI DAN PEMBAHASAN

Pada bab ini akan memaparkan hasil implementasi sistem yang telah dikembangkan, meliputi; 1. Hasil implementasi model deep learning, 2. Proses training dan evaluasi model, 3. Hasil optimalisasi untuk perangkat mobile.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan dikaji hasil penelitian dan pengujian perangkat lunak dengan metode tertentu yang digunakan dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini akan dikaji untuk kesimpulan dan saran dari penelitian yang telah ditulis yang diambil dari hasil penelitian yang telah dilakukan.

1.8 Kesimpulan

Sistem klasifikasi genre film yang akurat semakin dibutuhkan seiring dengan pertumbuhan industri perfilman nasional. Penelitian ini berfokus pada penerapan

model *IndoBERT* untuk mengklasifikasikan genre film berdasarkan sinopsis berbahasa Indonesia, dengan tujuan meningkatkan akurasi, efisiensi, dan konsistensi dalam proses klasifikasi.

Hasil penelitian diharapkan dapat berkontribusi dalam pengembangan metode klasifikasi teks berbahasa Indonesia serta menjadi dasar bagi sistem rekomendasi film berbasis teks. Untuk menjaga fokus, penelitian ini dibatasi pada teks sinopsis film dan tidak mencakup aspek lain seperti sentiment analysis atau entity recognition.

DAFTAR PUSTAKA

- Badan Perfilman Indonesia. (2025, 7 Januari). Wajah Perfilman Nasional Di Hari Film Nasional. https://www.bpi.or.id/artikel-27-wajah_perfilman_nasional_di_hari_film_nasional.html
- FilmIndonesia.Or.Id. (2024). Database Film Indonesia. <https://filmindonesia.or.id/> (Diakses: 24 November 2024)
- The Movie Database (Tmdb). (2024). The Movie Database. <https://www.themoviedb.org/> (Diakses: 26 November 2024)
- Hermansyah, D. D. (2024). Imdb Indonesian Movies. Kaggle. <https://www.kaggle.com/dionisiusdh/imdb-indonesian-movies> (Diakses: 26 November 2024)
- Christiyanto, A. (2024). Imdb Synopsis Indonesian Movies. Kaggle. <https://www.kaggle.com/antoniuscs/imdb-synopsis-indonesian-movies> (Diakses: 26 November 2024)
- Altman, R. (1999). *Film/Genre*. British Film Institute.
- Bordwell, D., & Thompson, K. (2019). *Film Art: An Introduction* (12th Ed.). Mcgraw-Hill Education.
- Jenkins, H. (2006). *Convergence Culture: Where Old And New Media Collide*. New York University Press.
- Akbar, J., Fahmi, H., & Murniati, W. (2025). Multi Label Klasifikasi Genre Film Berdasarkan Sinopsis Menggunakan Metode Long Short-Term Memory (Lstm). *Jurnal Manajemen Informatika & Sistem Informasi (Misi)*, 8(1). <https://doi.org/10.36595/misi.v5i2>
- Amelia, R., & Santoso, D. B. (2023). Prediksi Genre Film Dengan Klasifikasi Multi Kelas Sinopsis Menggunakan Jaringan Lstm Film Genre Prediction With Multi-Class Classification Synopsis Using Lstm Network. *Journal Of Information Technology And Computer Science (IntecomS)*, 6(2).

- Cahyawijaya, S., Winata, G. I., Wilie, B., Vincentio, K., Li, X., Kuncoro, A.,
Ruder, S., Lim, Z. Y., Bahar, S., Khodra, M. L., Purwarianti, A., & Fung, P.
(2021). *Indonlg: Benchmark And Resources For Evaluating Indonesian
Natural Language Generation*. [Http://Arxiv.Org/Abs/2104.08200](http://arxiv.org/abs/2104.08200)
- Feldman, Ronen., & Sanger, James. (2007). *The Text Mining Handbook :
Advanced Approaches In Analyzing Unstructured Data*. Cambridge
University Press.
- Hidayat, W. A., & Nastiti, V. R. S. (2024). Perbandingan Kinerja Pre-Trained
Indobert-Base Dan Indobert-Lite Pada Klasifikasi Sentimen Ulasan Tiktok
Tokopedia Seller Center Dengan Model Indobert. *Jsii (Jurnal Sistem
Informasi)*, 11(2), 13–20. [Https://Doi.Org/10.30656/Jsii.V11i2.9168](https://doi.org/10.30656/Jsii.V11i2.9168)
- Hong, C., Ghosh, R., & Srinivasan, S. (2016). *Dealing With Class Imbalance
Using Thresholding*. [Http://Arxiv.Org/Abs/1607.02705](http://arxiv.org/abs/1607.02705)
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020a). *Indolem And Indobert: A
Benchmark Dataset And Pre-Trained Language Model For Indonesian Nlp*.
[Http://Arxiv.Org/Abs/2011.00677](http://arxiv.org/abs/2011.00677)
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020b). *Indolem And Indobert: A
Benchmark Dataset And Pre-Trained Language Model For Indonesian Nlp*.
[Http://Arxiv.Org/Abs/2011.00677](http://arxiv.org/abs/2011.00677)
- Louppe, G., Pedregosa, F., Grisel, O., Buitinck, L., Blondel, M., Müller, A. C.,
Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R.,
Vanderplas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). *Api Design For*

Machine Learning Software: Experiences From The Scikit-Learn Project.

<https://www.researchgate.net/publication/256326897>

Madana, Y. P. (2021). *Perbandingan Model Klasifikasi Berdasarkan Input Teks (Sinopsis) Menggunakan Support Vector Machine Dan Convolutiunal Neural Network Terhadap Klasifikasi Genre Film.*

Muslimah, N., Indriati, & Cahya Wihandika, R. (2019). *Klasifikasi Film Berdasarkan Sinopsis Dengan Menggunakan Improved K-Nearest Neighbor (K-Nn)* (Vol. 3, Issue 1). <http://j-ptiik.ub.ac.id>

Nabila, H. M. (2020). *Klasifikasi Multilabel Genre Film Menggunakan Transformasi Dan Multilabel K-Nearest Neighbor (Ml-Knn).*

Nam, J., Loza Mencía, E., Kim, H. J., & Fürnkranz, J. (2017). *Maximizing Subset Accuracy With Recurrent Neural Networks In Multi-Label Classification.*

Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier Chains For Multi-Label Classification. *Machine Learning*, 85(3), 333–359.
<https://doi.org/10.1007/s10994-011-5256-5>

Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis Of Performance Measures For Classification Tasks. *Information Processing And Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>

Sorower, M. S. (2016). *A Literature Survey On Algorithms For Multi-Label Learning.* <https://www.researchgate.net/publication/266888594>

- Syafarina, G. A., & Zaenuddin. (2023). *Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi*. 7, 2023.
<https://doi.org/10.47002/Metik.V7i2.680>
- Tsoumakas, G., & Katakis, I. (2007). Multi-Label Classification: An Overview. *International Journal Of Data Warehousing And Mining*, 3(3), 1–13.
<https://doi.org/10.4018/Jdwm.2007070101>
- Wongso, W., Setiawan, D. S., Limcorn, S., & Joyoadikusumo, A. (2024). *Nusabert: Teaching Indobert To Be Multilingual And Multicultural*.
<http://arxiv.org/abs/2403.01817>
- Yang, H., Jin, Y., Li, Z., Wang, D.-B., Miao, L., Geng, X., & Zhang, M.-L. (2021). *Learning From Noisy Labels Via Dynamic Loss Thresholding*.
<http://arxiv.org/abs/2104.02570>
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent Trends In Deep Learning Based Natural Language Processing [Review Article]. In *IEEE Computational Intelligence Magazine* (Vol. 13, Issue 3, Pp. 55–75). Institute Of Electrical And Electronics Engineers Inc.
<https://doi.org/10.1109/MCI.2018.2840738>