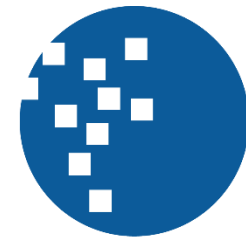


**ANALISIS PERBANDINGAN *LARGE LANGUAGE MODEL*
SEBAGAI *FEATURE EXTRACTION* DALAM DETEKSI
SARKASME MENGGUNAKAN ALGORITMA KLASIFIKASI**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Skripsi

Darrel Nathaniel Sabera

00000055439

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**ANALISIS PERBANDINGAN *LARGE LANGUAGE MODEL*
SEBAGAI *FEATURE EXTRACTION* DALAM DETEKSI
SARKASME MENGGUNAKAN ALGORITMA KLASIFIKASI**



SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer (S. Kom.)

Darrel Nathaniel Sabera

00000055439

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Darrel Nathaniel Sabera

Nomor Induk Mahasiswa : 00000055439

Program Studi : Sistem Informasi

Skripsi dengan judul:

“Analisis Perbandingan *Large Language Model* Sebagai *Feature Extraction* dalam Deteksi Sarkasme Menggunakan Algoritma Klasifikasi” Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 30 May 2025



Darrel Nathaniel Sabera

HALAMAN PENGESAHAN

Skripsi dengan judul

**ANALISIS PERBANDINGAN *LARGE LANGUAGE MODEL* SEBAGAI
FEATURE EXTRACTION DALAM DETEKSI SARKASME MENGGUNAKAN
ALGORITMA KLASIFIKASI**

Oleh

Nama : Darrel Nathaniel Sabera
NIM : 00000055439
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Selasa, 10 Juni 2025

Pukul 09.00 s.d 10.00 dan dinyatakan

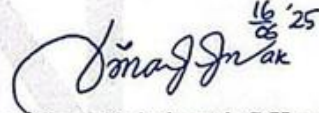
LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang


Dr. Irmawati, S.Kom., M.M.S.I.
0805097703/081431

Pembimbing


Dinar Ajeng Kristiyanti, S.Kom, M.Kom.
0330128801/079159

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.
03130588001/051314

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Darrel Nathaniel Sabera

NIM : 00000055439

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : Analisis Perbandingan *Large Language Model* Sebagai *Feature Extraction* dalam Deteksi Sarkasme Menggunakan Algoritma Klasifikasi

Menyatakan dengan sesungguhnya bahwa saya bersedia* (**pilih salah satu**):

- ☐ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☒ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 30 May 2025



(Darrel Nathaniel Sabera)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

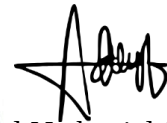
KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala Rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul "*Analisis Perbandingan Large Language Model Sebagai Feature Extraction dalam Deteksi Sarkasme Menggunakan Algoritma Klasifikasi*" dengan baik dan lancar. Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada program studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara. Saya menyadari bahwa terselesaikannya Tugas Akhir ini tidak lepas dari dukungan, bimbingan, dan doa dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

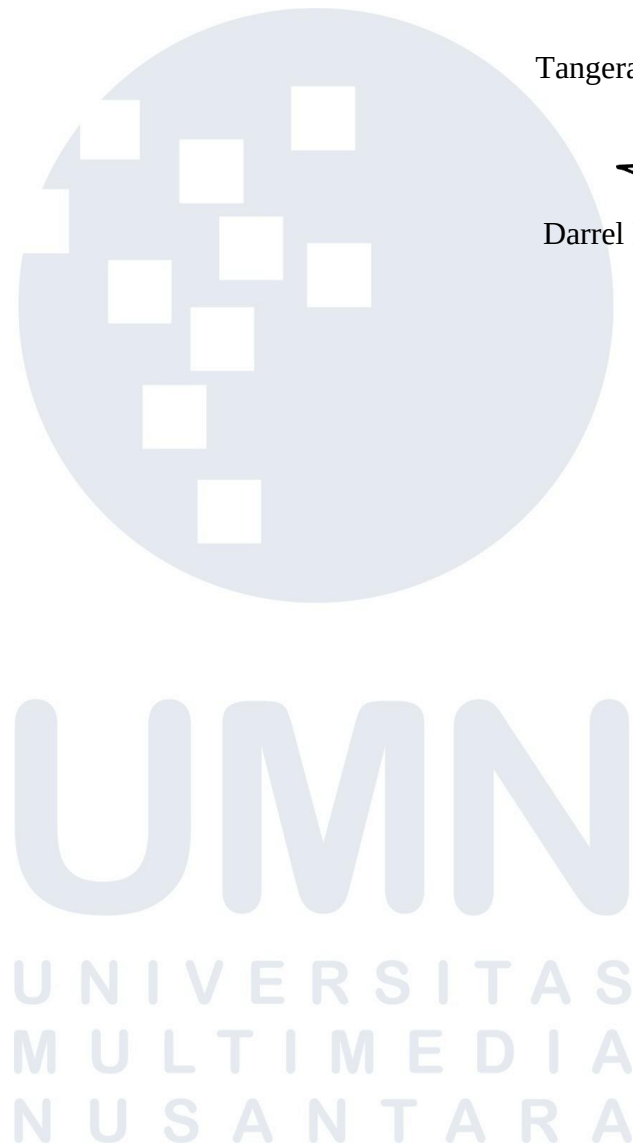
1. Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Dinar Ajeng Kristiyanti, S.Kom, M.Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Semua teman-teman seperjuangan di perkuliahan, yang telah menjadi bagian penting dalam perjalanan akademik ini. Terima kasih atas kebersamaan, semangat, diskusi, kerja sama, dan kenangan indah selama masa kuliah.

Saya menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat saya harapkan demi perbaikan di masa mendatang. Besar harapan saya, karya ini dapat memberikan manfaat, baik secara akademis maupun praktis, serta menjadi referensi yang berarti bagi para pembaca dan peneliti yang tertarik pada bidang Sistem Informasi.

Tangerang, 30 Mei 2025



Darrel Nathaniel Sabera



ANALISIS PERBANDINGAN *LARGE LANGUAGE MODEL* SEBAGAI *FEATURE EXTRACTION* DALAM DETEKSI SARKASME MENGGUNAKAN ALGORITMA KLASIFIKASI

Darrel Nathaniel Sabera

ABSTRAK

Deteksi sarkasme merupakan tugas penting dalam *NLP* (*Natural Language Processing*) karena ekspresi sarkasme dapat mendistorsi analisis sentimen dan menyesatkan sistem *decision machine*. Urgensi penelitian ini terletak pada keterbatasan metode tradisional seperti *Bag of Words* (*BoW*) dan *TF-IDF* yang gagal menangkap pemahaman kontekstual yang mendalam, sementara teknik *word embedding* seperti *Word2Vec* dan *GloVe* memberikan peningkatan, namun masih kesulitan dalam memahami makna kalimat secara utuh. *LLM* (*Large Language Model*) seperti *BERT* dan *RoBERTa* telah mentransformasi *NLP* dengan menangkap representasi kata secara kontekstual, sehingga lebih efektif untuk deteksi sarkasme. Penelitian ini menyajikan analisis komparatif berbagai metode ekstraksi fitur seperti *Word2Vec*, *GloVe*, *BERT*, dan *RoBERTa* yang dikombinasikan dengan algoritma klasifikasi seperti *Support Vector Machine* (*SVM*), *XGBoost*, dan *Random Forest*. Studi ini menggunakan *framework Knowledge Discovery in Database* (*KDD*) yang mencakup *data selection*, *preprocessing*, *transformation*, *modelling*, dan *evaluation*. Dataset yang digunakan terdiri dari judul berita yang diberi label sarkastik atau non-sarkastik. *Principal Component Analysis* digunakan untuk reduksi dimensi dengan mengeliminasi fitur-fitur yang redundan. Hasil penelitian menunjukkan bahwa kombinasi *RoBERTa-SVM* mencapai akurasi tertinggi sebesar 88,00%, yang menunjukkan keunggulan model berbasis *transformer* dibandingkan teknik *embedding* tradisional. Penelitian ini menyimpulkan bahwa integrasi *embedding* kontekstual dan seleksi fitur meningkatkan performa deteksi sarkasme sekaligus menjaga efisiensi komputasi. Namun, model ini masih menghadapi tantangan dalam mengidentifikasi sarkasme implisit karena tidak adanya petunjuk linguistik yang eksplisit, yang menjadi keterbatasan dari penelitian ini.

Kata kunci: *Feature Extraction, Feature Selection, Large Language Model, RoBERTa-Support Vector Machine, Sarcasm Detection*

COMPARATIVE ANALYSIS OF LARGE LANGUAGE MODEL AS FEATURE EXTRACTION METHOD IN SARCASM DETECTION USING CLASSIFICATION

ALGORITHMS

Darrel Nathaniel Sabera

ABSTRACT (English)

Sarcasm Detection is an important task in Natural Language Processing (NLP) because sarcasm expression can distort sentiment analysis and mislead automated decision making system. The urgency of this research lies in the limitations of traditional methods such as Bag of Words (BoW) and TF-IDF, which fail to capture deep contextual understanding, while word embedding techniques like Word2Vec and Glove have improvement, difficulties remained in comprehending full sentence meanings. Large Language Model (LLM) such as BERT and RoBERTa have Transformed NLP by capturing contextual word representation, making them more effective for sarcasm detection. This study presents a comparative analysis of various feature extraction method like Word2Vec, Glove, BERT, and RoBERTa combined with classification algorithms such as Support Vector Machine (SVM), XGBoost, and Random Forest. This study uses the Knowledge Discovery in Database (KDD) Framework which includes data selection, preprocessing, transformation, modeling, and evaluation. The dataset consists of news headlines labeled as sarcastic or non sarcastic, Principal Component Analysis used for dimensionality reduction by eliminating redundant features. The result show that the RoBERTa-SVM combination achieves the highest accuracy of 88.00%, indicating the superiority of transformer based models over traditional embedding techniques. This study concludes that integrating contextual embeddings and feature selection improves sarcasm detection performance while maintaining computational efficiency. However, the model still faces challenges in identifying implicit sarcasm due the absence of explicit linguistic cues which represent a limitation of this research.

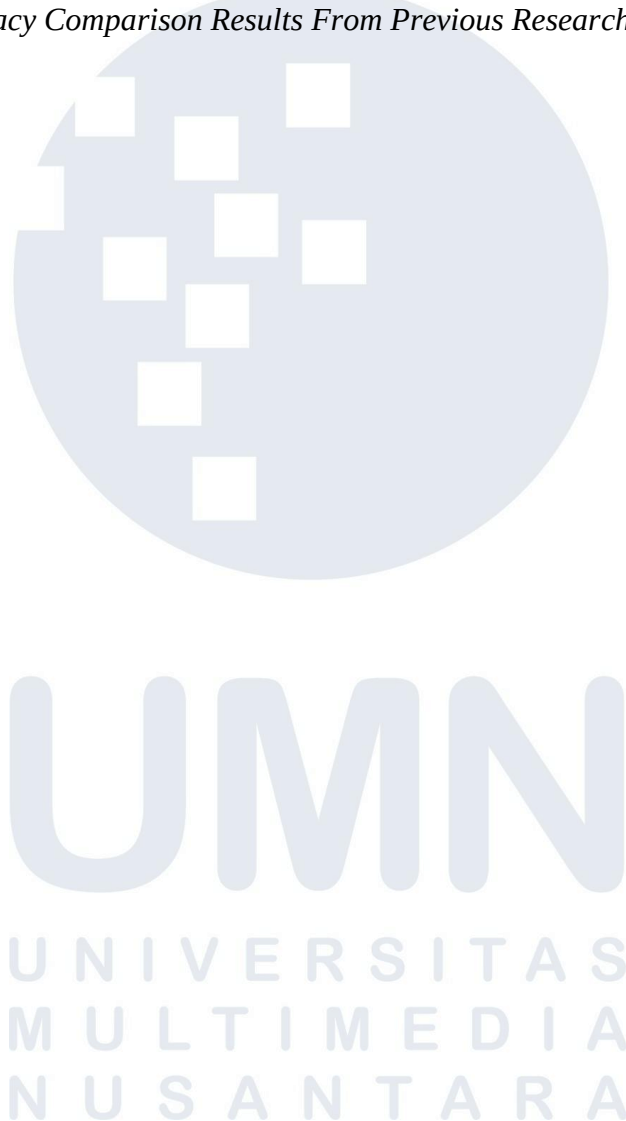
Keywords: Feature Extraction, Feature Selection, Large Language Model, RoBERTa–Support Vector Machine, Sarcasm Detection

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
I. INTRODUCTION.....	1
II. RELATED WORKS	2
III. RESEARCH METHOD	2
3.1 Data Selection.....	2
3.2 Data Preprocessing and Transformation	3
3.3 Feature Extraction using Large Language Model	3
3.4 Feature Selection.....	3
3.5 Classification	4
3.6 Evaluation.....	4
IV. RESULT AND DISCUSSION	4
4.1 Comparison of Large Language Model Results without Feature Selection	4
4.2 Comparison of Large Language Model Results using Feature Selection .	5
4.3 Discussion.....	5
V. CONCLUSION.....	6
ACKNOWLEDGMENT.....	6
REFERENCES	6
LAMPIRAN.....	8

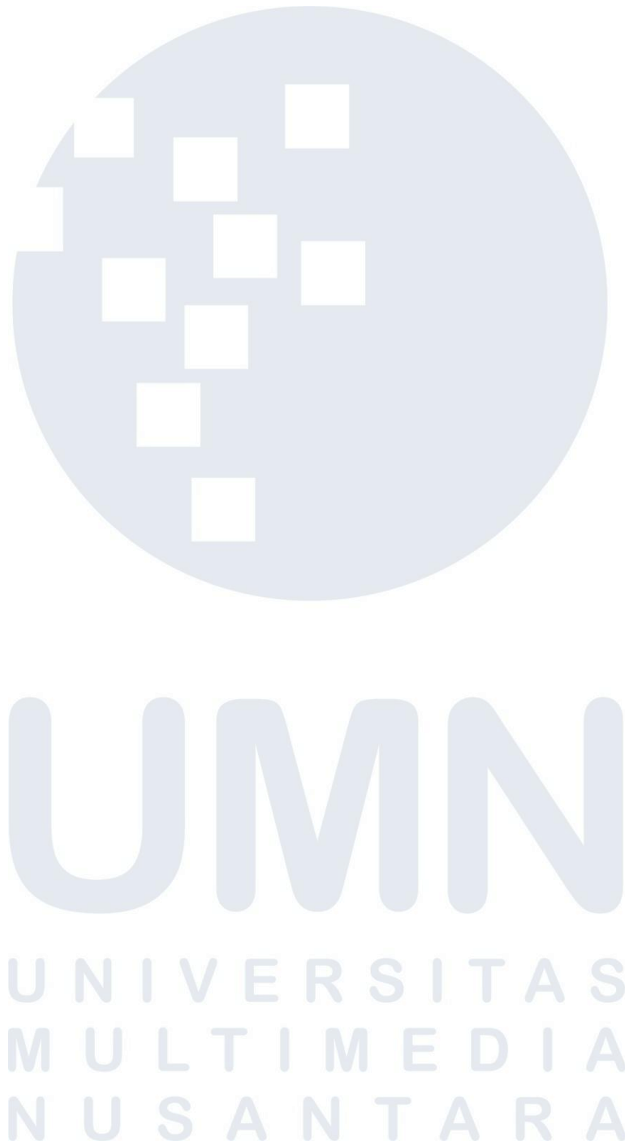
DAFTAR TABEL

<i>Table I. Example of Sarcasm Data</i>	<i>3</i>
<i>Table II. Comparison of Large Language Model Results without Feature Selection</i>	<i>5</i>
<i>Table III. Comparison of Large Language Model Results using Feature Selection</i>	<i>5</i>
<i>Table IV. Accuracy Comparison Results From Previous Research</i>	<i>5</i>



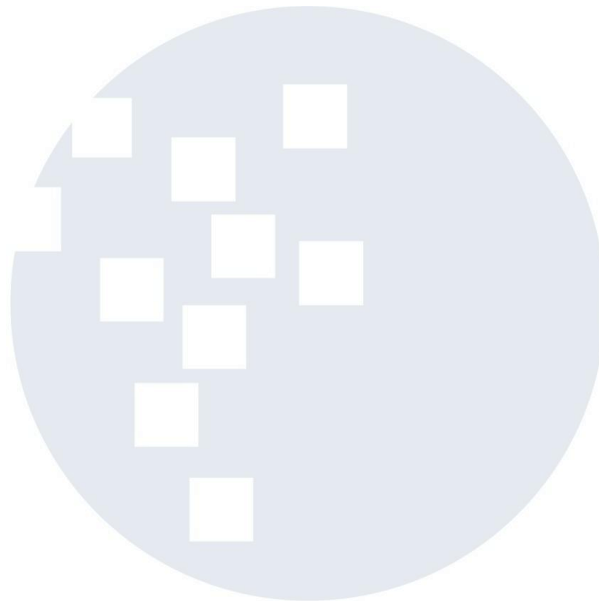
DAFTAR GAMBAR

<i>Fig. 1 Research Workflow</i>	2
<i>Fig. 2. Confusion Matrix RoBERTa-SVM Model Based using PCA</i>	5



DAFTAR LAMPIRAN

Lampiran A Form Konsultasi Bimbingan.....	8
Lampiran B <i>Letter of Acceptance (LoA)</i>	10
Lampiran C Hasil Turnitin	11
Lampiran D <i>Dataset</i>	19
Lampiran E <i>Conference Certificate</i>	20



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA