

**PENERAPAN MODEL INDOBERT DALAM DETEKSI SARKASME
MENGUNAKAN RANDOM FOREST CLASSIFIER
PADA ANALISIS SENTIMEN**

TESIS

Oleh:

**SABRINA ADELA BR SIBARANI
NIM. 221231038**



**UNIVERSITAS
MIKROSKIL**

**PROGRAM STUDI S-2 MAGISTER TEKNOLOGI INFORMASI
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2024**

**IMPLEMENTATION OF INDOBERT IN SARCASM DETECTION
USING RANDOM FOREST CLASSIFIER TOWARDS
SENTIMENT ANALYSIS**

THESIS

By:

**SABRINA ADELA BR SIBARANI
NIM. 221231038**



**UNIVERSITAS
MIKROSKIL**

**MAJOR OF S-2 MAGISTER INFORMATION TECHNOLOGY
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2024**

LEMBARAN PENGESAHAN

**PENERAPAN MODEL INDOBERT DALAM DETEKSI SARKASME
MENGUNAKAN RANDOM FOREST CLASSIFIER
PADA ANALISIS SENTIMEN**

TESIS

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Magister
Program Studi S-2 Teknologi Informasi

Oleh:

SABRINA ADELA BR SIBARANI
NIM. 221231038

Disetujui Oleh:

Dosen Pembimbing I,

UNIVERSITAS
MIKROSKIL

Dr. Ronsen Purba, M.Sc.

Medan, 2024

Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-2 Teknologi Informasi,

Ir. Erwin Setiawan Panjaitan, MMSI., Ph.D.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa program studi S-2 Teknologi Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Sabrina Adela Br Sibarani
NIM : 221231038

Saya telah melaksanakan penelitian dan penulisan Tesis dengan judul sebagai berikut:
Judul Tesis : Penerapan Model Indobert dalam Deteksi Sarkasme Menggunakan *Random Forest Classifier* pada Analisis Sentimen

Sehubungan dengan Tesis tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*) atas Tesis saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tesis saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tesis saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 11 Juli 2024

Saya yang membuat pernyataan,



Sabrina Adela Br Sibarani

PENERAPAN MODEL INDOBERT DALAM DETEKSI SARKASME MENGUNAKAN RANDOM FOREST CLASSIFIER PADA ANALISIS SENTIMEN

ABSTRAK

Sarkasme adalah bentuk sindiran kompleks yang menciptakan perbedaan antara arti harfiah kata dan maksud sebenarnya. Penelitian ini mengevaluasi deteksi sarkasme dalam komentar media sosial menggunakan Random Forest Classifier (RFC) yang dipadukan dengan IndoBERT. Pengujian dilakukan dengan TfidfVectorizer dan 10-Fold Cross Validation. Tanpa IndoBERT, model RFC mencapai rata-rata nilai akurasi, recall, dan F1-score mencapai 78.83%, serta nilai presisi sebesar 79.01%. Dengan integrasi IndoBERT, performa model meningkat signifikan dengan rata-rata akurasi, presisi, recall, dan F1score masing-masing 84.22%, 85.05%, 84.22%, dan 84.15%. Pengujian 5-Fold Cross Validation pada model RFC dengan IndoBERT menghasilkan akurasi tertinggi sebesar 97.24%, presisi 97.39%, recall, dan F1 score sebesar 97.24%. Penelitian ini menyarankan eksplorasi model deep learning lain, pencarian hyperparameter yang lebih mendalam, serta penambahan fitur dan perluasan dataset untuk meningkatkan akurasi deteksi sarkasme.

Kata kunci: deteksi sarkasme, random forest classifier, IndoBERT, TfidfVectorizer, crossvalidation

Abstract

Sarcasm is a complex form of irony that creates a discrepancy between the literal meaning of words and their intended message. This study assesses sarcasm detection in social media comments using a combination of Random Forest Classifier (RFC) and IndoBERT. Testing was conducted with TfidfVectorizer and 10-Fold Cross Validation. Without IndoBERT, the RFC model achieved average accuracy, precision, recall, and F1-score of 73.29%, 73.15%, 73.77%, and 72.87%, respectively. Integrating IndoBERT significantly improved performance, with average accuracy, precision, recall, and F1-score of 84.21%, 85.05%, 84.21%, and 84.15%. The 5-fold cross-validation on the RFC model with IndoBERT achieved the highest accuracy of 97.24%, precision of 97.39%, recall, and F1 score of 97.24%. The study suggests exploring alternative deep learning models, more rigorous hyperparameter tuning, and expanding features and datasets to further enhance sarcasm detection accuracy.

Keywords: sarcasm detection, Random Forest, IndoBERT, TfidfVectorizer, 10-Fold Cross Validation

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Bapa yang Esa, Tuhan Yesus Kristus atas berkat dan kasih karunia-Nya, penulis dapat menyelesaikan Tesis yang berjudul “PENERAPAN MODEL INDOBERT DALAM DETEKSI SARKASME MENGGUNAKAN RANDOM FOREST CLASSIFIER PADA ANALISIS SENTIMEN”.

Tesis ini dibuat untuk melengkapi persyaratan kurikulum pada Program Studi S-2 Teknologi Informasi, Universitas Mikroskil Medan.

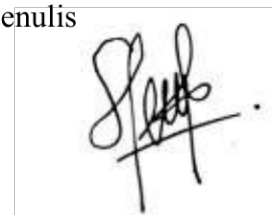
Penulis juga mengucapkan banyak Terimakasih kepada:

1. Bapak Dr. Ronsen Purba, M.Sc., selaku Dosen Pembimbing I yang sudah mendukung dan mau sangat sabar membimbing serta memberikan saran kepada penulis dalam penyusunan Tesis ini.
2. Bapak Muhammad Fermi Pasha, B.Sc., M.Sc., PhD, selaku Dosen Pendamping Pembimbing yang telah membantu dalam penyusunan Tesis ini.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil.
4. Bapak Ng Poi Wong, S.Kom., M.T.I., selaku Dekan Fakultas Informatika Universitas Mikroskil.
5. Bapak Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D., selaku Ketua Program Studi S-2 Teknologi Informasi Universitas Mikroskil.
6. Orang tua yang senantiasa memberikan dukungan dan doa yang selalu dipanjatkan kepada penulis.
7. Bapak dan Ibu Dosen Universitas Mikroskil yang telah memberikan dukungan dan semangat selama proses penulisan tesis ini.
8. Abang Ricky, teman-teman penulis (Juliana, Bobby, Fandi) dan semua pihak yang sudah memberikan dukungan kepada penulis untuk menyelesaikan penulisan tesis ini.

Penulis menyadari bahwa Tesis ini masih terdapat banyak kekurangan karena keterbatasan kemampuan dari penulis. Oleh karena itu, kritik dan saran dari pembaca sangat diharapkan untuk penulisan yang lebih baik di masa yang akan datang dan semoga Tesis ini dapat memberikan manfaat bagi pembaca.

Medan, 11 Juli 2024

Penulis

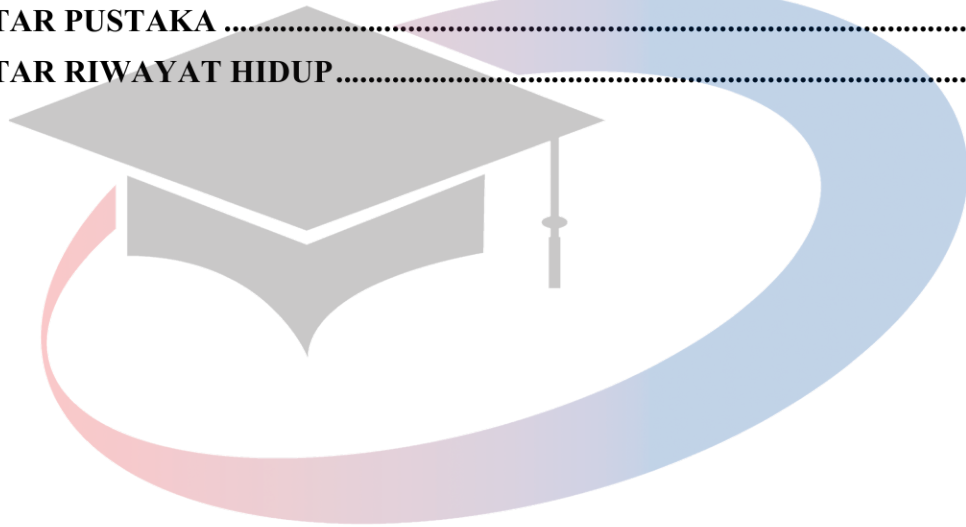


Sabrina Adela Br Sibarani

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan.....	2
1.4 Manfaat.....	3
1.5 Ruang Lingkup	3
BAB II KAJIAN LITERATUR	4
2.1 Tinjauan Pustaka	4
2.1.1 Media Sosial	4
2.1.2 Analisis Sentimen.....	4
2.1.3 Sarkasme dalam Teks	7
2.1.4 BERT	10
2.1.5 IndoBERT.....	11
2.1.6 <i>Random Forest Classifier</i>	12
2.1.7 <i>Cross Validation</i>	15
2.1.8 <i>Confusion Matrix</i>	15
2.2 Penelitian Terdahulu.....	18
2.3 Kerangka Konsep	20
BAB III METODOLOGI PENELITIAN	21
3.1 Analisis Masalah	21
3.2 Rancangan Metodologi Penelitian.....	21
3.2.1 Data yang Digunakan	22
3.2.2 Analisis Sentimen	23
3.2.3 Deteksi Sarkasme dengan Klasifikasi <i>Random Forest</i>	25
3.2.4 Penerapan Model IndoBERT	27
3.2.5 Evaluasi Model	27
3.3 Alat-alat Penelitian	28
3.4 Jadwal Penelitian	28
BAB IV HASIL DAN PEMBAHASAN	29
4.1 Hasil.....	29
4.1.1 Persiapan Data	29

4.1.2 Analisis Sentimen	30
4.1.3 Pemodelan Deteksi Sarkasme.....	34
4.1.4 Pelatihan Model Deteksi dengan RFC	39
4.1.5 Penerapan Model IndoBERT pada Model Deteksi	40
4.2 Pembahasan	42
4.2.1 Evaluasi Model Deteksi Menggunakan RFC	42
4.2.2 Evaluasi Penerapan Model IndoBERT pada Model RFC	43
BAB V PENUTUP	44
5.1 Kesimpulan.....	44
5.2 Saran.....	44
DAFTAR PUSTAKA	45
DAFTAR RIWAYAT HIDUP.....	48



UNIVERSITAS MIKROSKIL

DAFTAR GAMBAR

Gambar 2.1	Macam-macam Media Sosial	4
Gambar 2.2	Preprocessing Analisis Sentimen Menggunakan Machine Learning [20].....	5
Gambar 2.3	Contoh Sarkasme	7
Gambar 2.4	Pre-Training dan Fine-Tuning pada BERT [28].....	11
Gambar 2.5	Proses Embedding pada IndoBERT [11].....	12
Gambar 2.6	Decision Tree Model [30].....	13
Gambar 2.7	Random Forest Classifier	14
Gambar 2.8	10-fold Cross Validation [31].....	15
Gambar 2.9	Ilustrasi Confusion Matrix [32]	16
Gambar 2.10	Kerangka Konsep	20
Gambar 3.1	Flowchart Metode Penelitian.....	21
Gambar 3.2	Video Youtube dan Contoh Komentar pada Channel Najwa Shihab.....	22
Gambar 3.3	Flowchart Analisis Sentimen.....	23
Gambar 3.4	Flowchart Deteksi Sarkasme dengan RFC	25
Gambar 3.5	Proses Deteksi Sarkasme dengan Menerapkan Model IndoBERT.....	27
Gambar 4.1	Hasil Pemindahan Data Teks dari Kolom “Reply” ke Kolom “Comment”	31
Gambar 4.2	Hasil Penghapusan Kolom.....	31
Gambar 4.3	Hasil Teks setelah Dilakukan Casefolding.....	32
Gambar 4.4	Hasil Penghapusan Karakter dan Tokenisasi.....	32
Gambar 4.5	Hasil Stemming	33
Gambar 4.6	Hasil Vektorisasi.....	33
Gambar 4.7	Hasil Analisis Sentimen.....	34
Gambar 4.8	Proses Punctuation Relate.....	36
Gambar 4.9	Menghitung Jumlah Skor Sarkasme	38
Gambar 4.10	Hasil Pelabelan Sarkasme Berdasarkan Ekstraksi Fitur	38
Gambar 4.11	Proses Pembagian Data.....	39
Gambar 4.12	Vektorisasi dengan TF-IDF	39
Gambar 4.13	Proses Pelatihan Model dan Prediksi.....	40
Gambar 4.14	Preprocessing pada IndoBERT	40
Gambar 4.15	Tokenisasi	41
Gambar 4.16	Proses Ekstraksi Fitur Menggunakan IndoBERT.....	41
Gambar 4.17	Proses Pemodelan RFC	41

DAFTAR TABEL

Tabel 2.1 Beberapa Penelitian Deteksi sarkasme[26]	9
Tabel 2.2 Confusion Matrix [32]	16
Tabel 3.1 Jadwal Penelitian	28
Tabel 4.1 Hasil Evaluasi Model RFC	42
Tabel 4.2 Hasil Evaluasi Model Deteksi dengan IndoBERT	43



UNIVERSITAS
MIKROSKIL