

KOMPARASI ALGORITMA KLASIFIKASI *FINE-TUNING* INDOBERT DAN MULTINOMIAL NAÏVE BAYES UNTUK ANALISIS SENTIMEN ULASAN APLIKASI THREADS PADA GOOGLE PLAY STORE



**OLEH
KOMANG HOKKY ARYASTA
NIM 2015091028**

**PROGRAM STUDI SISTEM INFORMASI
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA**

2024



KOMPARASI ALGORITMA KLASIFIKASI *FINE-TUNING* INDOBERT DAN MULTINOMIAL NAÏVE BAYES UNTUK ANALISIS SENTIMEN ULASAN APLIKASI THREADS PADA GOOGLE PLAY STORE

SKRIPSI

**Diajukan kepada
Universitas Pendidikan Ganesha
Untuk Memenuhi Salah Satu Persyaratan dalam Menyelesaikan
Program Sarjana Komputer**

**Oleh
Komang Hokky Aryasta
NIM 2015091028**

**PROGRAM STUDI SISTEM INFORMASI
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA**

2024

SKRIPSI

**DIAJUKAN UNTUK MELENGKAPI TUGAS
DAN MEMENUHI SYARAT-SYARAT UNTUK
MENCAPAI GELAR SARJANA KOMPUTER**

Menyetujui

Pembimbing I,



I Gede Mahendra Darmawiguna, S.Kom., M.Sc.
NIP. 198501042010121004

Pembimbing II,



Gede Surya Mahendra, S.Pd., M.Kom.
NIP. 199003132022031009

Skripsi oleh Komang Hokky Aryasta
telah dipertahankan di depan dewan penguji
Pada tanggal 25 Januari 2024

Dewan Penguji,



I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T.
NIP. 198907112020122004

(Ketua)



I Made Dendi Maysanjaya, S.Pd., M.Eng.
NIP. 199005152019031008

(Anggota)



I Gede Mahendra Darmawiguna, S.Kom., M.Sc.
NIP. 198501042010121004

(Anggota)



Gede Surya Mahendra, S.Pd., M.Kom.
NIP. 199003132022031009

(Anggota)

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha
guna memenuhi syarat-syarat untuk mencapai gelar Sarjana Komputer

Pada:

Hari : Kamis
Tanggal : 25 Januari 2024

Mengetahui,

Ketua Ujian,



Made Windu Antara Kesiman, S.T., M.Sc., Ph.D.
NIP. 198211112008121001

Sekretaris Ujian,

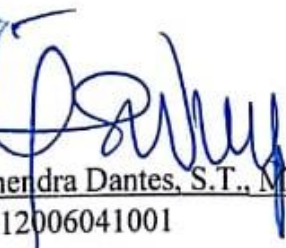


I Made Dendi Maysanjaya, S.Pd., M.Eng.
NIP. 199005152019031008

Mengesahkan,

Dekan Fakultas Teknik dan Kejuruan




Dr. Kadek Rihendra Dantes, S.T., M.T.
NIP. 197912012006041001

PERNYATAAN

Dengan ini saya menyatakan bahwa karya tulis yang berjudul “KOMPARASI ALGORITMA KLASIFIKASI FINE-TUNING · INDOBERT DAN MULTINOMIAL NAÏVE BAYES UNTUK ANALISIS SENTIMEN ULASAN APLIKASI THREADS PADA GOOGLE PLAY STORE” beserta seluruh isinya adalah benar-benar karya sendiri dan saya tidak melakukan penjiplakan dan pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran atas etika keilmuan dalam karya saya ini atau ada klaim terhadap keaslian karya saya ini.

Singaraja, 25 Januari 2024

Yang membuat pernyataan,



Komang Hokky Aryasta

NIM. 2015091028

KATA PERSEMBAHAN

“Om Swastyastu”

“Om Deva Suksma Parama Achintya ya namah swaha

Sarwa karya prasidhantam”

:

Om Sang Hyang Widhi Wasa dalam wujud Parama Achintya yang maha gaib nan maha karya, hanya atas anugrahMu-lah maka pekerjaan ini berhasil dengan baik.

Dengan mengucapkan puji syukur atas berkat rahmat-Nya, skripsi ini saya dedikasikan kepada :

**KEDUA ORANG TUA TERCINTA
(Komang Yudi Aryasuta & Ni Made Sumartini)**

Yang telah menyayangi, membimbing, dan mendukung penulis dengan penuh kasih sayang. Terima kasih atas semangat, motivasi, dukungan, dan doa yang telah diberikan dalam setiap langkah yang penulis ambil. Terima kasih sudah menjadi obor penyemangat penulis dalam menyelesaikan jenjang pendidikan ini.

SAUDARA TERSAYANG

Yang selalu menghibur dan menyemangati penulis ketika sedang merasa lelah saat proses pengerjaan skripsi ini dan semoga kita semua menjadi anak yang membanggakan untuk kedua orang tua kita.

REKAN-REKAN SEPERJUANGAN

I Gusti Ngurah Agung Pernata, Gede Anggi Widiantara, dan Putu Agisna Fransisco Sara yang selalu menjadi teman seperjuangan dan tempat penulis berkeluh kesah selama penulis menempuh pendidikan sarjana.

MOTTO



PRAKATA

Puji syukur penyusun panjatkan ke hadapan Tuhan Yang Maha Esa karena berkat rahmat-Nya lah, penulis dapat menyelesaikan skripsi yang berjudul “Komparasi Algoritma Klasifikasi Fine-Tuning IndoBERT dan Multinomial Naïve Bayes Untuk Analisis Sentimen Ulasan Aplikasi Threads Pada Google Play Store”. Skripsi ini disusun guna memenuhi persyaratan mencapai gelar sarjana komputer pada Universitas Pendidikan Ganesha.

Dalam menyelesaikan skripsi ini, penulis banyak mendapat bantuan baik berupa moral maupun material dari berbagai pihak. Untuk itu, dalam kesempatan ini, penulis mengucapkan terima kasih kepada :

1. Bapak Dr. Kadek Rihendra Dantes, S.T., M.T., selaku Dekan Fakultas Teknik dan Kejuruan atas motivasi dan fasilitas yang diberikan sehingga penulis bisa menyelesaikan studi sesuai dengan rencana .
2. Bapak Putu Hendra Suputra, S.Kom., M.Cs., selaku Ketua Jurusan Teknik Informatika atas motivasi dan fasilitas yang diberikan dalam penyelesaian skripsi ini.
3. Bapak I Gede Mahendra Darmawiguna, S.Kom., M.Sc., selaku Pembimbing I yang telah memberikan bimbingan, arahan, petunjuk, dan motivasi kepada penulis dalam penyelesaian skripsi ini.
4. Bapak Gede Surya Mahendra, S.Pd., M.Kom., selaku Pembimbing II yang telah memberikan bimbingan, arahan, petunjuk, dan motivasi kepada penulis dalam penyelesaian skripsi ini.
5. Seluruh Dosen Jurusan Teknik Informatika/Program Studi Sistem Informasi yang telah berbagi ilmu dan pengalamannya selama kuliah semoga semua yang peneliti pelajari di kampus dapat bermanfaat bagi masyarakat luas.
6. Secara khusus orang tua, saudara, serta sahabat yang telah mendoakan dan memberikan dukungan yang sangat berarti kepada peneliti dalam menyelesaikan skripsi ini.
7. Rekan mahasiswa Program Studi Sistem Informasi yang telah memberikan motivasi kepada penulis dalam penyelesaian skripsi ini.

8. Semua pihak yang tidak bisa disebutkan satu per satu yang telah membantu penyusunan skripsi ini

Penulis menyadari sepenuhnya bahwa apa yang tersaji dalam skripsi ini masih jauh dari kata sempurna karena keterbatasan kemampuan yang penulis miliki. Untuk itu demi kesempurnaan skripsi ini, penulis mengharapkan segala kritik maupun saran yang sifatnya membangun dari berbagai pihak. Penulis berharap skripsi ini dapat bermanfaat dan berguna bagi kita semua.



Singaraja, 25 Januari 2024

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN LOGO	ii
HALAMAN JUDUL.....	iii
LEMBAR PERSETUJUAN PEMBIMBING	iv
LEMBAR PERSETUJUAN DOSEN PENGUJI.....	v
LEMBAR PERSETUJUAN DAN PENGESAHAN PANITIA UJIAN	vi
PERNYATAAN KEASLIAN TULISAN.....	vii
KATA PERSEMBAHAN.....	viii
MOTTO.....	ix
PRAKATA.....	x
ABSTRAK	xii
<i>ABSTRACT</i>	xiii
DAFTAR ISI.....	xiv
DAFTAR TABEL.....	xvii
DAFTAR GAMBAR	xix
DAFTAR LAMPIRAN	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Batasan Masalah Penelitian.....	5
1.5. Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA	6
2.1. Penelitian Terdahulu	6

2.1.1.	Penelitian Terkait Analisis Sentimen pada Google Play Store	6
2.1.2.	Penelitian Terkait dengan IndoBERT	6
2.1.3.	Penelitian Terkait dengan Multinomial Naïve Bayes	7
2.2.	Dasar Teori	8
2.2.1.	Media Sosial	9
2.2.2.	Threads.....	10
2.2.3.	Google Play Store	11
2.2.4.	Data Mining	12
2.2.5.	Text Mining.....	14
2.2.6.	Sastrawi.....	15
2.2.7.	Klasifikasi	16
2.2.8.	Analisis Sentimen	16
2.2.9.	BERT	17
2.2.10.	<i>Fine-tuning</i> BERT.....	21
2.2.11.	IndoBERT	22
2.2.12.	Naïve Bayes Classifier.....	24
2.2.13.	<i>K-Fold Cross Validation</i>	26
2.2.14.	Confusion Matrix.....	27
BAB III METODOLOGI PENELITIAN.....		29
3.1.	Tahapan Penelitian.....	29
3.2.	Metode Pengumpulan Data	31
3.3.	Pengolahan dan Analisis Data	31
3.4.	Implementasi Model.....	37
3.5.	Evaluasi Model.....	41
BAB IV HASIL DAN PEMBAHASAN.....		43
4.1.	Hasil.....	43

4.1.1.	Pengumpulan Dataset	43
4.1.2.	Data Selection	44
4.1.3.	Data Labeling.....	44
4.1.4.	Data Preprocessing	45
4.1.5.	Hasil Pengolahan Data.....	48
4.1.6.	Implementasi Model	48
4.2.	Pembahasan	65
4.2.1.	Perbandingan Performa Konfigurasi <i>Fine-Tuning</i> IndoBERT	65
4.2.2.	Analisis Hasil Komparasi Model.....	65
4.2.3.	Analisis Hasil Sentimen Ulasan.....	66
BAB V PENUTUP.....		70
5.1.	Kesimpulan.....	70
5.2.	Saran.....	70
DAFTAR PUSTAKA.....		71
LAMPIRAN.....		77



DAFTAR TABEL

Tabel 2.1 Confusion Matrix Multiclass.....	27
Tabel 3.1 Sampel <i>Dataset</i>	32
Tabel 3.2 Contoh <i>Labeling Data</i>	33
Tabel 3.3 Contoh <i>Case Folding</i>	34
Tabel 3.4 Contoh <i>Cleansing</i>	34
Tabel 3.5 Contoh <i>Tokenizing</i>	35
Tabel 3.6 Contoh <i>Stopword Removal</i>	35
Tabel 3.7 Contoh <i>Stemming</i>	36
Tabel 3.8 Contoh <i>Normalization</i>	36
Tabel 3.9 Proses Tokenisasi Teks pada IndoBERT	38
Tabel 3.10 Ilustrasi Implementasi <i>Bag of Words</i>	39
Tabel 3.11 Contoh Data Perhitungan Algoritma Multinomial Naïve Bayes	40
Tabel 4.1 Hasil <i>Scraping</i>	43
Tabel 4.2 Hasil Data <i>Selection</i>	44
Tabel 4.3 Hasil Data <i>Labeling</i>	44
Tabel 4.4 Hasil <i>Case Folding</i>	45
Tabel 4.5 Hasil <i>Cleansing</i>	45
Tabel 4.6 Hasil <i>Tokenizing</i>	46
Tabel 4.7 Hasil <i>Stopword Removal</i>	46
Tabel 4.8 Hasil <i>Stemming</i>	47
Tabel 4.9 Hasil <i>Normalization</i>	47
Tabel 4.10 Hasil Pengolahan Data	48
Tabel 4.11 Distribusi Label	48
Tabel 4.12 Spesifikasi Google Colaboratory	49
Tabel 4.13 Hasil <i>Word Embedding</i> IndoBERT.....	49
Tabel 4.14 Konfigurasi <i>Hyperparameter</i> 1	51
Tabel 4.15 Konfigurasi <i>Hyperparameter</i> 2	53
Tabel 4.16 Konfigurasi <i>Hyperparameter</i> 3	55
Tabel 4.17 Konfigurasi <i>Hyperparameter</i> 4	57
Tabel 4.18 Konfigurasi <i>Hyperparameter</i> 5	59
Tabel 4.19 Konfigurasi <i>Hyperparameter</i> 6	61

Tabel 4.17 Hasil <i>Bag of Words</i>	63
Tabel 4.18 Perbandingan Enam Konfigurasi <i>Hyperparameter</i>	65
Tabel 4.19 Perbandingan Performa Kedua Model	65
Tabel 4.20 Data Ulasan Terprediksi Benar	69



DAFTAR GAMBAR

Gambar 2.1 Tampilan Aplikasi Threads	10
Gambar 2.2 Proses Knowledge Discovery In Database (KDD)	13
Gambar 2.3 Arsitektur Transformer	17
Gambar 2.4 Arsitektur BERT dengan Arsitektur Model Lainnya.....	18
Gambar 2.5 Representasi Input BERT	19
Gambar 2.6 Tahapan Pre-training dan Fine-tuning BERT.....	20
Gambar 2.7 Proses Masked LM.....	21
Gambar 2.8 Ilustrasi <i>Fine-tuning</i> BERT	22
Gambar 2.9 Ilustrasi Arsitektur IndoBERT.....	24
Gambar 2.10 Contoh 10-fold Cross Validation.....	26
Gambar 3.1 Diagram Alir Tahapan Penelitian	29
Gambar 3.2 Tahapan Pengolahan dan Analisis Data.....	31
Gambar 3.3 Tahapan <i>Pre-processing Data</i>	34
Gambar 3.4 Kamus IndoBERT	37
Gambar 3.5 Ilustrasi Proses <i>Fine-Tuning</i> IndoBERT	39
Gambar 3.6 Contoh Simulasi <i>Confusion Matrix</i>	42
Gambar 4.1 Grafik Akurasi Per Fold IndoBERT Konfigurasi 1	51
Gambar 4.2 Confusion Matrix IndoBERT Konfigurasi 1	52
Gambar 4.3 Grafik Akurasi Per Fold IndoBERT Konfigurasi 2	53
Gambar 4.4 Confusion Matrix IndoBERT Konfigurasi 2	54
Gambar 4.5 Grafik Akurasi Per Fold IndoBERT Konfigurasi 3	55
Gambar 4.6 Confusion Matrix IndoBERT Konfigurasi 3	56
Gambar 4.7 Grafik Akurasi Per Fold IndoBERT Konfigurasi 4.....	57
Gambar 4.8 Confusion Matrix IndoBERT Konfigurasi 4.....	58
Gambar 4.9 Grafik Akurasi Per Fold IndoBERT Konfigurasi 5	59
Gambar 4.10 Confusion Matrix IndoBERT Konfigurasi 5	60
Gambar 4.11 Grafik Akurasi Per Fold IndoBERT Konfigurasi 6	61
Gambar 4.12 Confusion Matrix IndoBERT Konfigurasi 6	62
Gambar 4.13 Grafik Akurasi Per Fold Multinomial Naïve Bayes	63
Gambar 4.14 Confusion Matrix Multinomial Naïve Bayes	64
Gambar 4.15 <i>WordCloud</i> Sentimen Positif	66

Gambar 4.16 <i>WordCloud</i> Sentimen Netral.....	66
Gambar 4.17 <i>WordCloud</i> Sentimen Negatif.....	67
Gambar 4.18 Distribusi Sentimen dengan <i>fine-tuning</i> IndoBERT.....	68



DAFTAR LAMPIRAN

Lampiran 1. Riwayat Hidup	77
Lampiran 2. Kode <i>Scraping Data</i>	77
Lampiran 3. Kode <i>Data Selection</i>	78
Lampiran 4. Kode <i>Data Labeling</i>	78
Lampiran 5. Kode <i>Case Folding</i>	78
Lampiran 6. Kode <i>Cleansing</i>	78
Lampiran 7. Kode <i>Tokenizing</i>	79
Lampiran 8. Kode <i>Stopword Removal</i>	79
Lampiran 9. Kode <i>Stemming</i>	79
Lampiran 10. Kode <i>Normalization</i>	80
Lampiran 11. Kode <i>Fine-Tuning IndoBERT</i>	80
Lampiran 12. Kode Multinomial Naïve Bayes	86

