

**PERBANDINGAN EKSTRAKSI FITUR *TF-IDF* DAN  
*WORD2VEC* PADA ANALISIS SENTIMEN  
MENGUNAKAN MODEL *FINE-TUNING*  
INDOBERT UNTUK ULASAN GAME LOKAL DI  
STEAM**



**OLEH  
I GUSTI NGURAH AGUNG PERNATA  
NIM 2015091033**

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

**2024**



**PERBANDINGAN EKSTRAKSI FITUR *TF-IDF* DAN  
*WORD2VEC* PADA ANALISIS SENTIMEN  
MENGUNAKAN MODEL *FINE-TUNING*  
INDOBERT UNTUK ULASAN GAME LOKAL DI  
STEAM**



**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 I Gusti Ngurah Agung Pernata  
telah dipertahankan di depan dewan penguji  
Pada tanggal 6 Februari 2024

Dewan Penguji,



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

(Ketua)



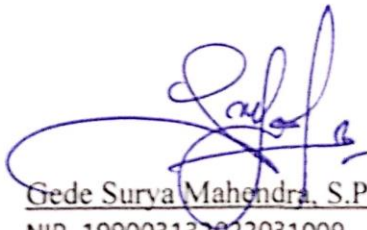
I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T.  
NIP. 198907112020122004

(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 : Selasa  
Tanggal : 6 Februari 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. Kodek Rihendra Dantes, S.T., M.T.  
NIP. 197912012006041001

## PERNYATAAN

Dengan ini saya menyatakan bahwa karya tulis yang berjudul “PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN WORD2VEC PADA ANALISIS SENTIMEN MENGGUNAKAN MODEL FINE-TUNING INDOBERT UNTUK ULASAN GAME LOKAL DI STEAM” 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, 6 Februari 2024

Yang membuat pernyataan,



I Gusti Ngurah Agung Pernata

NIM. 2015091033

## 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

**(I Gusti Putu Ngurah Upeksa (alm) & Ni Luh Ariwati)**

Yang telah memberi dukungan, menyayangi, membimbing, serta selalu memberikan segalanya dengan penuh kasih sayang. Terima kasih atas semangat, motivasi dan doa yang telah diberikan dalam setiap langkah yang penulis ambil mulai dari sebelum masuk perguruan tinggi, hingga jenjang saat ini. Terima kasih sudah menjadi obor penyemangat penulis dalam menyelesaikan jenjang pendidikan ini.

### SAUDARA DAN SELURUH KELUARGA TERSAYANG

Yang selalu memberikan motivasi 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

Komang Hokky Aryasta, I Gusti Made Rai Laksana, dan seluruh rekan lainnya telah membantu meberikan dukungan baik secara langsung maupun tidak langsung, sehingga memberikan semangat dan motivasi lebih kepada penulis untuk menyelesaikan program sarjana.



## MOTTO

“

***Everyone Has Their Own Story***

”



## PRAKATA

Puji syukur penyusun panjatkan ke hadapan Tuhan Yang Maha Esa karena berkat rahat-Nya lah, penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Ekstraksi Fitur TF-IDF Dan Word2Vec Pada Analisis Sentimen Menggunakan Model Fine-Tuning Indobert Untuk Ulasan Game Lokal Di Steam”. 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, 6 Februari 2024

Penulis



## DAFTAR ISI

HALAMAN SAMPUL .....	i
HALAMAN LOGO .....	ii
HALAMAN JUDUL.....	iii
HALAMAN PENGESAHAN DOSEN PEMBIMBING.....	iv
HALAMAN PENGESAHAN DOSEN PENGUJI .....	v
PERNYATAAN KEASLIAN TULISAN .....	vii
KATA PERSEMBAHAN.....	viii
MOTTO .....	ix
PRAKATA.....	x
ABSTRAK.....	xii
<i>ABSTRACT</i> .....	xiii
DAFTAR ISI.....	xiv
DAFTAR GAMBAR .....	xvii
DAFTAR TABEL.....	xix
DAFTAR LAMPIRAN.....	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah .....	5
1.3. Tujuan Penelitian.....	5
1.4. Batasan Masalah Penelitian.....	5
1.5. Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Penelitian Terdahulu.....	7
2.1.1. Penelitian Terkait Analisis Sentimen pada Steam .....	7

2.1.2.	Penelitian Terkait Perbandingan Ekstraksi Fitur TF-IDF dan Word2Vec.....	8
2.1.3.	Penelitian Terkait Implementasi IndoBERT.....	9
2.2.	Dasar Teori.....	11
2.2.1.	Steam.....	11
2.2.2.	<i>Data Mining</i> .....	13
2.2.3.	Text Mining.....	14
2.2.4.	Sastrawi.....	16
2.2.5.	Analisis Sentimen.....	16
2.2.6.	BERT.....	17
2.2.7.	IndoBERT.....	19
2.2.8.	Ekstraksi Fitur.....	20
2.2.9.	<i>K-Fold Cross-Validation</i> .....	22
2.2.10.	<i>Confusion Matrix</i> .....	23
<b>BAB III METODOLOGI PENELITIAN.....</b>		<b>26</b>
3.1.	Tahapan Penelitian.....	26
3.2.	Metode Pengumpulan Data.....	28
3.3.	Pengelolaan dan Analisis Data.....	28
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>42</b>
4.1.	Hasil.....	42
4.1.1.	Pengumpulan Data.....	42
4.1.2.	Data Selection.....	44
4.1.3.	Data Labelling.....	45
4.1.4.	Data Preprocessing.....	45
4.1.5.	Hasil Pengolahan Data.....	48
4.1.6.	Implementasi Model.....	49

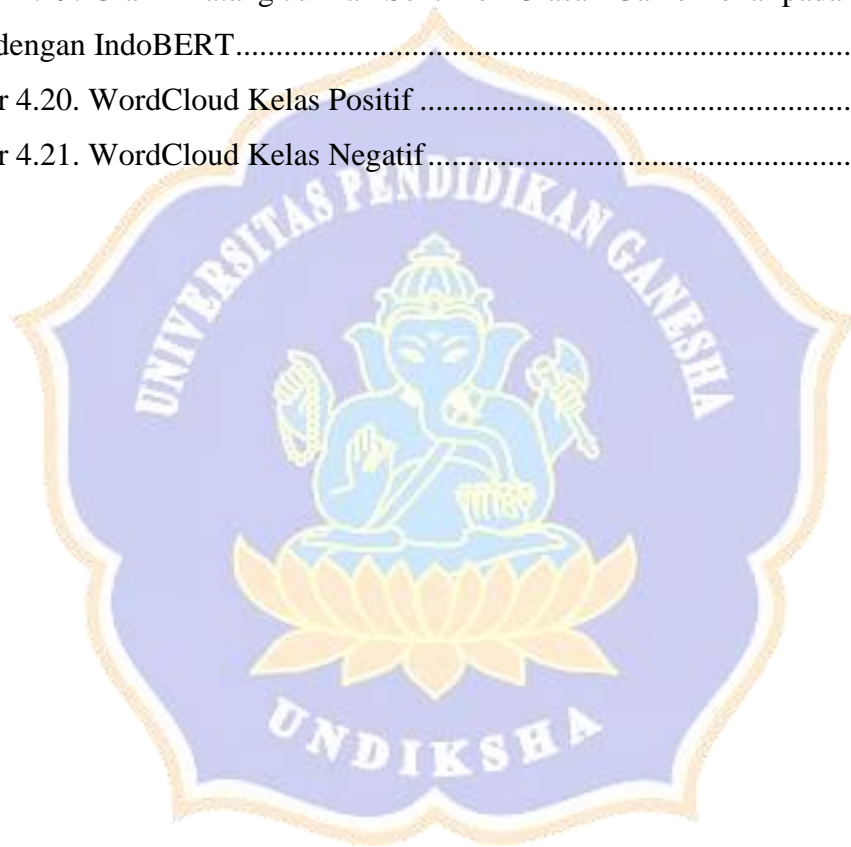
4.2. Pembahasan .....	68
4.2.1. Perbandingan Performa Konfigurasi.....	68
4.2.2. Analisis Hasil Komparasi Algoritma .....	70
4.2.3. Analisis Hasil Sentimen pada Ulasan Game Lokal di Steam .....	71
BAB V PENUTUP.....	76
5.1. Kesimpulan.....	76
5.2. Saran .....	76
DAFTAR PUSTAKA .....	78
LAMPIRAN.....	86



## DAFTAR GAMBAR

Gambar 2.1. Tampilan Awal Steam.....	12
Gambar 2.2. Tampilan Steam Review .....	12
Gambar 2.3. Tahapan Data Mining.....	14
Gambar 2.4. Prosedur Pre-training dan Fine-Tuning pada BERT.....	18
Gambar 2.5. Pendekatan CBOW dan Skip-gram.....	22
Gambar 2.6. Ilustrasi K-Fold Cross-Validation .....	23
Gambar 3.1. Tahapan Penelitian .....	26
Gambar 3.2. Alur Pengolahan dan Analisis Data .....	28
Gambar 3.3. Tahapan Text Preprocessing .....	34
Gambar 4.1. Grafik Akurasi Setiap Fold IndoBERT untuk Konfigurasi Pertama	51
Gambar 4.2. Confusion Matrix model IndoBERT Konfigurasi Pertama.....	52
Gambar 4.3. Grafik Akurasi Setiap Fold IndoBERT Konfigurasi Kedua .....	53
Gambar 4.4. Confusion Matrix model IndoBERT untuk Konfigurasi Kedua.....	54
Gambar 4.5. Grafik Akurasi Setiap Fold IndoBERT untuk Konfigurasi Ketiga..	55
Gambar 4.6. Confusion Matrix model IndoBERT Konfigurasi Ketiga .....	56
Gambar 4.7. Grafik Akurasi Setiap Fold IndoBERT dengan Ekstraksi Fitur TF-IDF Konfigurasi Pertama.....	58
Gambar 4.8. Confusion Matrix model IndoBERT dengan Ekstraksi Fitur TF-IDF Konfigurasi Pertama.....	58
Gambar 4.9. Grafik Akurasi Setiap Fold IndoBERT dengan Ekstraksi Fitur TF-IDF Konfigurasi Kedua .....	59
Gambar 4.10. Confusion Matrix model IndoBERT dengan Ekstraksi Fitur TF-IDF Konfigurasi Kedua .....	60
Gambar 4.11. Grafik Akurasi Setiap Fold IndoBERT dengan Ekstraksi Fitur TF- IDF Konfigurasi Ketiga.....	61
Gambar 4.12. Confusion Matrix model IndoBERT dengan Ekstraksi Fitur TF-IDF Konfigurasi Ketiga .....	61
Gambar 4.13. Grafik Akurasi Setiap Fold IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Pertama .....	63
Gambar 4.14. Confusion Matrix model IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Pertama .....	64

Gambar 4.15. Grafik Akurasi Setiap Fold IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Kedua.....	65
Gambar 4.16. Confusion Matrix model IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Kedua.....	65
Gambar 4.17. Grafik Akurasi Setiap Fold IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Ketiga .....	66
Gambar 4.18. Confusion Matrix model IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Ketiga .....	67
Gambar 4.19. Grafik Batang Jumlah Sentimen Ulasan Game Lokal pada Platform Steam dengan IndoBERT.....	72
Gambar 4.20. WordCloud Kelas Positif .....	74
Gambar 4.21. WordCloud Kelas Negatif.....	74





## DAFTAR TABEL

Tabel 2.1. Visualisasi Confusion Matrix.....	23
Tabel 3.1. Contoh Dataset Ulasan Game Lokal Pada Steam .....	29
Tabel 3.2. Contoh data yang telah diberikan label.....	31
Tabel 3.3. Contoh hasil tahapan case folding .....	34
Tabel 3.4. Contoh hasil tahapan cleansing.....	35
Tabel 3.5. Contoh hasil tahapan tokenizing .....	36
Tabel 3.6. Contoh hasil tahapan stopword removal.....	36
Tabel 3.7. Contoh hasil tahapan stemming .....	37
Tabel 3.8. Contoh hasil tahapan normalization.....	38
Tabel 3.9 Contoh Hasil Confusion Matrix dengan 10-Fold Cross Validation.....	41
Tabel 3.10 Contoh Hasil Pengukuran Kinerja .....	41
Tabel 4.1. Daftar Video Games Lokal dan Jumlah Ulasan.....	42
Tabel 4.2. Hasil Scrapping Ulasan Steam.....	43
Tabel 4.3. Hasil Data Selection.....	44
Tabel 4.4. Hasil Data Labelling .....	45
Tabel 4.5. Hasil Case Folding .....	45
Tabel 4.6. Hasil Cleansing .....	46
Tabel 4.7. Hasil Tokenizing .....	46
Tabel 4.8. Hasil Stopword Removal .....	47
Tabel 4.9. Hasil Stemming .....	47
Tabel 4.10. Hasil Normalization .....	48
Tabel 4.11. Hasil Pengolahan Data.....	48
Tabel 4.12. Distribusi Label.....	49
Tabel 4.13. Hasil Word Embedding IndoBERT .....	50
Tabel 4.14. Konfigurasi Hyperparameter IndoBERT yang pertama .....	51
Tabel 4.15. Classification Report Model IndoBERT Konfigurasi Pertama.....	52
Tabel 4.16. Performa rata-rata model IndoBERT Konfigurasi Kedua .....	53
Tabel 4.17. Classification Report Model IndoBERT Konfigurasi Kedua .....	54
Tabel 4.18. Performa rata-rata model IndoBERT Konfigurasi Ketiga .....	55
Tabel 4.19. Classification Report Model IndoBERT Konfigurasi Ketiga.....	56
Tabel 4.20. Matrix TF-IDF .....	57

Tabel 4.21. Classification Report Model IndoBERT dengan Ekstraksi Fitur TF-IDF Konfigurasi Pertama.....	59
Tabel 4.22. Classification Report Model IndoBERT dengan Ekstraksi Fitur TF-IDF Konfigurasi Kedua .....	60
Tabel 4.23. Classification Report Model IndoBERT dengan Ekstraksi Fitur TF-IDF Konfigurasi Ketiga .....	62
Tabel 4.24. Hasil Vektor Word2Vec.....	62
Tabel 4.25. Classification Report Model IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Pertama .....	64
Tabel 4.26. Classification Report Model IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Kedua.....	66
Tabel 4.27. Classification Report Model IndoBERT dengan Ekstraksi Fitur Word2Vec Konfigurasi Ketiga .....	67
Tabel 4.28. Perbandingan Performa Kofigurasi Hyperparameter IndoBERT .....	68
Tabel 4.29. Perbandingan Performa Kofigurasi Hyperparameter IndoBERT dengan Ekstraksi Fitur TF-IDF.....	69
Tabel 4.30. Perbandingan Performa Kofigurasi Hyperparameter IndoBERT dengan Ekstraksi Fitur Word2Vec.....	70
Tabel 4.31. Perbandingan Performa Masing-masing Metode.....	70
Tabel 4.32. Data Ulasan Terprediksi Benar .....	73

## DAFTAR LAMPIRAN

Lampiran 1. Riwayat Hidup .....	86
Lampiran 2 : Kode Scrapping Ulasan Steam .....	86
Lampiran 3 : Kode Data Selection .....	92
Lampiran 4 : Kode Data Labelling.....	92
Lampiran 5 : Kode Case Folding .....	93
Lampiran 6 : Kode Cleansing .....	93
Lampiran 7 : Kode Tokenizing .....	95
Lampiran 8 : Kode Stopword Removal .....	95
Lampiran 9 : Kode Stemming .....	96
Lampiran 10 : Kode Normalization .....	96
Lampiran 11 : Kode Fine-tuning IndoBERT .....	96
Lampiran 12 : Kode Fine-tuning IndoBERT dengan Ekstraksi Fitur TF-IDF ...	105
Lampiran 13 : Kode Fine-tuning IndoBERT dengan Ekstraksi Fitur Word2Vec .....	109

