

**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI ACCESS
BY KAI MENGGUNAKAN METODE WORD2VEC DAN SUPPORT
VECTOR MACHINE**

SKRIPSI



Disusun Oleh:

DITHA LOZERA DEVI

20082010143

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
S U R A B A Y A
2024**

SKRIPSI

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI ACCESS BY KAI MENGGUNAKAN METODE WORD2VEC DAN SUPPORT VECTOR MACHINE

Disusun Oleh:
DITHA LOZERA DEVI
20082010143

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi
Program Studi Sistem Informasi Fakultas Ilmu Komputer
Universitas Pembangunan Nasional “Veteran” Jawa Timur
pada Tanggal 12 Juni 2024

Pembimbing :

1.



Amalia Anjani Arifiyanti, S.Kom., M.Kom.
NIP. 19920812 201803 2 001

2.



Seftin Fitri Ana Wati, S.Kom., M.Kom.
NPT. 2/12199 10 320267

Tim Penguji :

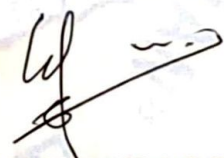
1.


Agung Brastama Putra, S.Kom., M.Kom.
NIP. 19851124 202121 1 003

2.


Siti Mukaromah, S.Kom., M.Kom.
NIP. 19810704 2021212 011

3.


Abdul Rezha Efrat Najaf, S.Kom., M.Kom.
NIP. 19940029 2022031 008

Mengetahui,
Dekan Fakultas Ilmu Komputer
Universitas Pembangunan Nasional “Veteran” Jawa Timur


Prof. Dr. Ir. Novirina Hendrasarie, M.T.
NIP. 19681126 199403 2 001

LEMBAR PENGESAHAN

**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI ACCESS BY
KAI MENGGUNAKAN METODE WORD2VEC DAN SUPPORT VECTOR
MACHINE**

Disusun Oleh:
DITHA LOZERA DEVI
20082010143


**Telah disetujui mengikuti Ujian Negara Lisan Gelombang Juni
Periode 2024 pada Tanggal 12 Juni 2024**

Menyetujui,

Dosen Pembimbing 1

Dosen Pembimbing 2


Amalia Anjani Arifiyanti, S.Kom., M.Kom.
NIP. 19920812 201803 2 001


Seftin Fitri Ana Wati, S.Kom., M.Kom.
NPT. 212199 10 320267

Mengetahui,
Ketua Program Studi Sistem Informasi
Fakultas Ilmu Komputer


Agung Brastama Putra, S.Kom., M.Kom.
NIP. 19851124 2021211 003



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
FAKULTAS ILMU KOMPUTER

KETERANGAN REVISI

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : Ditha Lozera Devi

NPM : 20082010143

Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 12 Juni 2024 dengan judul:

"ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI ACCESS BY KAI MENGGUNAKAN METODE WORD2VEC DAN SUPPORT VECTOR MACHINE"

Oleh karenanya mahasiswa tersebut di atas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diijinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, 24 Juni 2024

Dosen penguji yang memeriksa skripsi:

1. Agung Brastama Putra, S.Kom., M.T.

NIP. 19851124 202121 1 003

{  }

2. Siti Mukaromah, S. Kom., M.Kom.

NIP. 19810704 2021212 011

{  }

3. Abdul Rezha Efrat Najaf, S.Kom., M.Kom.

NIP. 19940029 2022031 008


{  }

Mengetahui,

Dosen Pembimbing 1

Dosen Pembimbing 2


Amalia Anjani Arifiyanti, S.Kom., M.Kom.
NIP. 19920812 201803 2 001


Seftin Fitri Ana Wati, S.Kom., M.Kom.
NPT. 212199 10 320267



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
FAKULTAS ILMU KOMPUTER**

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Ditha Lozera Devi

NPM : 20082010143

Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**“ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI ACCESS BY
KAI MENGGUNAKAN METODE WORD2VEC DAN SUPPORT VECTOR
MACHINE”**

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir / Penelitian orang lain dan juga bukan merupakan Produk / Hasil Karya yang saya beli dari orang lain.

Saya juga menyatakan bahwa Skripsi / Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institusi Pendidikan lain. Jika ternyata dikemudian hari pernyataan terbukti benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari.

Surabaya, 24 Juni 2024

Hormat Saya,



Ditha Lozera Devi

NPM. 20082010143

Judul : ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
ACCESS BY KAI MENGGUNAKAN METODE
WORD2VEC DAN SUPPORT VECTOR MACHINE

Pembimbing 1 : Amalia Anjani Arifiyanti, S.Kom., M.Kom

Pembimbing 2 : Seftin Fitri Ana Wati, S.Kom., M.,Kom

ABSTRAK

Tahun 2014, PT. Kereta Api Indonesia yang merupakan perusahaan Badan Usaha Milik Negara (BUMN) berinovasi dengan meluncurkan aplikasi KAI Access. Namun, beberapa ulasan pengguna yang menyatakan bahwa *loading* aplikasi lambat, batas waktu pemesanan tiket terlalu cepat, dan sering terjadi gangguan saat pemesanan tiket. Hingga akhirnya pada tanggal 10 Agustus 2023 PT. Kereta Api Indonesia melakukan peluncuran aplikasi Access by KAI sebagai bentuk *upgrade* dari aplikasi sebelumnya yaitu KAI Access.

Adanya ulasan yang diberikan pengguna untuk aplikasi, perlu dilakukan analisis sentimen untuk melihat bagaimana pendapat dan reaksi pengguna dalam menggunakan aplikasi Access by KAI. Skripsi ini menggunakan metode Word2vec sebagai metode ekstraksi fitur dan algoritma *Support Vector Machine* (SVM) untuk analisis sentimen pengguna. Metode Word2vec yang digunakan terdapat 2 model yaitu CBOW dan Skip-gram. Terdapat 4 kernel SVM yang digunakan yaitu kernel *linear*, kernel *polynomial*, kernel RBF, dan kernel *sigmoid*.

Hasil dari delapan skenario model klasifikasi menggunakan 4 kernel SVM dan 2 model Word2vec, dihasilkan satu skenario terbaik yang memiliki nilai akurasi dan nilai AUC (*Area Under the Curve*) tinggi. Skenario terbaik itu adalah skenario model yang menggunakan algoritma SVM kernel RBF dengan metode Skip-Gram ditambah metode *oversampling* SMOTE dihasilkan nilai akurasi 81% dan nilai AUC sebesar 0.81. Sehingga skenario tersebut divisualisasikan ke dalam bentuk *website*.

Kata Kunci: Klasifikasi, Analisis Sentimen, Word2vec, SVM, Access by KAI.

KATA PENGANTAR

Puji syukur saya panjatkan kehadiran Allah SWT. yang telah melimpahkan rahmat dan hidayah sehingga saya dapat menyelesaikan skripsi ini yang berjudul “Analisis Sentimen Ulasan Pengguna Aplikasi Access by Kai Menggunakan Metode Word2vec dan Support Vector Machine” sebagai salah satu syarat dalam menyelesaikan Program Studi Sistem Informasi di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam penyusunan skripsi ini, banyak pihak yang telah memberi bantuan dan dukungan sehingga skripsi ini dapat selesai tepat waktu. Oleh karena itu, penulis dengan penuh hormat mengucapkan terima kasih kepada:

1. Bapak Miyanto, Ibu Sulastri, kakak, dan adik saya selaku keluarga yang telah memberikan dukungan materi serta doa agar skripsi ini bisa selesai.
2. Ibu Amalia Anjani Arifiyanti, S.Kom., M.Kom selaku dosen pembimbing 1 yang senantiasa memberikan arahan dengan sabar dalam menyelesaikan skripsi ini.
3. Ibu Seftin Fitri Ana Wati, S.Kom., M.Kom selaku dosen pembimbing 2 yang juga senantiasa memberikan arahan dengan sabar dalam menyelesaikan skripsi ini.
4. Bapak Agung Brastama Putra, S.Kom., M.Kom selaku Koordinator Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Seluruh dosen Program Studi Sistem Informasi atas ilmu-ilmu yang telah diberikan selama masa perkuliahan.

6. Seluruh teman-teman Program Studi Sistem Informasi angkatan 2020 terutama Apriliana Pramesinta K. dan Safina Nur Rahmah yang telah membantu menjadi bagian pengukir cerita penulis selama masa perkuliahan.
7. Sahabat saya Dena Orintasari dan Yasmin Fajar Putri A. yang telah mendengarkan keluh kesah saya dan senantiasa memberikan semangat.

Semoga Tuhan memberikan balasan yang setimpal kepada semua pihak yang telah memberikan bantuan, bimbingan serta motivasi kepada penulis. Penulis menyadari bahwa Laporan Skripsi ini masih banyak kekurangan. Namun, penulis berharap semoga Laporan Skripsi ini dapat memberikan manfaat dalam perkembangan ilmu pengetahuan bagi pembaca.

Surabaya, Juni 2024

Penulis

DAFTAR ISI

LEMBAR PERNYATAAN	i
LEMBAR PENGESAHAN	ii
KETERANGAN REVISI.....	iii
SURAT PERNYATAAN.....	iv
ABSTRAK	v
KATA PENGANTAR.....	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan.....	5
1.5 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Dasar Teori	7
2.1.1 Analisis Sentimen	7
2.1.2 Access by KAI	7
2.1.3 Google Play Store	8
2.1.4 App Store	8
2.1.5 Teks Mining	8
2.1.6 SEMMA	9
2.1.7 <i>Word Embedding</i>	11
2.1.8 Word2Vec	12
2.1.9 <i>Support Vector Machine (SVM)</i>	13
2.1.10 Evaluasi Model.....	17
2.1.10.1 <i>Confusion Matrix</i>	17
2.1.10.2 <i>Accuracy</i>	18
2.1.10.3 <i>Precision</i>	19
2.1.10.4 <i>Recall</i>	19

2.1.10.5	F1-Score	19
2.1.10.6	ROC Curve	19
2.1.10.7	AUC	20
2.1.11	Python	20
2.1.12	SMOTE	21
2.2	Penelitian Terdahulu	22
2.2.1	Penelitian Terdahulu Hastari Utama Tahun 2022.....	22
2.2.2	Penelitian Terdahulu Gracia Rediena Tahun 2023	22
2.2.3	Penelitian Terdahulu Huda Mustakim Tahun 2022	23
2.2.4	Penelitian Terdahulu Kiki Ahmad Dasuki Tahun 2023	23
2.2.5	Penelitian Terdahulu Emi Suryati Tahun 2023.....	23
2.2.6	Penelitian Terdahulu Rona Guines Purnasiwi Tahun 2023	24
BAB III METODOLOGI PENELITIAN.....		25
3.1	Studi Literatur.....	25
3.2	Analisis Kebutuhan	25
3.2.1	Kebutuhan Perangkat Keras.....	25
3.2.2	Kebutuhan Perangkat Lunak	26
3.3	SEMMA	26
3.3.1	Sample	27
3.3.2	Explore	28
3.3.3	Modify	29
3.3.3.1	Pelabelan Data.....	29
3.3.3.2	Preprocessing Data.....	29
3.3.3.2.1	Cleansing	29
3.3.3.2.2	Case Folding.....	30
3.3.3.2.3	Stopword Removal	30
3.3.3.2.4	Tokenizing	31
3.3.3.2.5	Stemming.....	31
3.3.3.3	Ekstraksi Fitur (Word2Vec).....	32
3.3.3.4	Data Splitting.....	32
3.3.3.5	SMOTE	33
3.3.4	Model	33
3.3.4.1	Skenario Pertama.....	33
3.3.4.2	Skenario Kedua	34

3.3.4.3	Skenario Ketiga	34
3.3.4.4	Skenario Keempat	34
3.3.4.5	Skenario Kelima	35
3.3.4.6	Skenario Keenam	35
3.3.4.7	Skenario Ketujuh	35
3.3.4.8	Skenario Kedelapan	35
3.3.5	Assess	36
3.4	Pengembangan Sistem	36
3.5	Uji Validasi Sistem	36
BAB IV HASIL DAN PEMBAHASAN		37
4.1	SEMMA	37
4.1.1	<i>Sample</i>	37
4.1.2	<i>Explore</i>	39
4.1.3	<i>Modify</i>	42
4.1.3.1	Pelabelan Data	42
4.1.3.2	<i>Preprocessing Data</i>	43
4.1.3.2.1	<i>Cleansing</i>	44
4.1.3.2.2	<i>Case Folding</i>	45
4.1.3.2.3	<i>Stopword Removal</i>	47
4.1.3.2.4	<i>Tokenizing</i>	48
4.1.3.2.5	<i>Stemming</i>	50
4.1.3.3	Ekstraksi Fitur (Word2Vec)	51
4.1.3.4	<i>Data Splitting</i>	54
4.1.3.5	SMOTE	54
4.1.4	Model	56
4.1.4.1	SVM Kernel Linear	56
4.1.4.2	SVM Kernel Polynomial	57
4.1.4.3	SVM Kernel RBF	57
4.1.4.4	SVM Kernel Sigmoid	58
4.1.5	Assess	58
4.1.5.1	SVM Linear & CBOW	59
4.1.5.2	SVM Linear & Skip-Gram	60
4.1.5.3	SVM Polynomial & CBOW	62
4.1.5.4	SVM Polynomial & Skip-Gram	64

4.1.5.5	SVM RBF & CBOW.....	66
4.1.5.6	SVM RBF & Skip-Gram.....	68
4.1.5.7	SVM Sigmoid & CBOW.....	70
4.1.5.8	SVM Sigmoid & Skip-Gram.....	72
4.1.6	Pembahasan.....	74
4.2	Pengembangan Sistem.....	78
4.3	Uji Validasi Sistem.....	80
BAB V PENUTUP.....		81
5.1	Kesimpulan.....	81
5.2	Saran.....	82
DAFTAR PUSTAKA.....		83
LAMPIRAN.....		87

DAFTAR GAMBAR

Gambar 1.1. Ulasan dan Rating pada Google Play Store	1
Gambar 2.1. Model SEMMA.....	9
Gambar 2.2. Arsitektur Skip-Gram (Sivakumar dkk., 2020)	12
Gambar 2.3. Arsitektur CBOW (Sivakumar dkk., 2020).....	13
Gambar 2.4. Ilustrasi Linear SVM (Cervantes dkk., 2020)	14
Gambar 2.5. Ilustrasi Linear Kernel.....	15
Gambar 2.6. Ilustrasi Polynomial Kernel.....	16
Gambar 2.7. Ilustrasi RBF Kernel	16
Gambar 2.8. Ilustrasi Sigmoid Kernel	17
Gambar 3.1. Tahapan Penelitian	25
Gambar 3.2. Tahapan Model SEMMA	26
Gambar 3.3. Proses Pelabelan Data	29
Gambar 3.4. Proses Cleansing Data.....	29
Gambar 3.5. Proses Case Folding	30
Gambar 3.6. Proses Stopword Removal	30
Gambar 3.7. Proses Tokenizing.....	31
Gambar 3.8. Proses Stemming.....	31
Gambar 3.9. Proses Ekstraksi Fitur.....	32
Gambar 3.10. Proses SMOTE.....	33
Gambar 4.1. Source code scraping data dari Google Play Store	37
Gambar 4.2. Source code scraping data dari App Store.....	37
Gambar 4.3. Source code filter data.....	38
Gambar 4.4 Source code explore data.....	39
Gambar 4.5 Source code visualisasi histogram	40
Gambar 4.6 Grafik histogram jumlah rating ulasan.....	40
Gambar 4.7. Source code visualisasi violin plot.....	41
Gambar 4.8. Violin plot atribut rating.....	41
Gambar 4.9. Source code pelabelan data	42
Gambar 4.10. Source code cleansing data	44
Gambar 4.11. Source code case folding.....	46
Gambar 4.12. Source code stopword removal	47

Gambar 4.13. Source code tokenizing	49
Gambar 4.14. Source code stemming	50
Gambar 4.15. Source code CBOW	52
Gambar 4.16. Source code Skip-Gram	53
Gambar 4.17. Source code data splitting	54
Gambar 4.18. Source code SMOTE.....	55
Gambar 4.19. Source code SVM Linear	56
Gambar 4.20. Source code SVM polynomial	57
Gambar 4.21. Source code SVM RBF	57
Gambar 4.22. Source code SVM sigmoid.....	58
Gambar 4.23. Confusion matrix SVM Linear & CBOW.....	60
Gambar 4.24. Confusion matrix SVM Linear & Skip-Gram.....	62
Gambar 4.25. Confusion matrix SVM Polynomial & CBOW	64
Gambar 4.26. Confusion matrix SVM Polynomial & Skip-Gram.....	66
Gambar 4.27. Confusion matrix SVM RBF & CBOW	68
Gambar 4.28. Confusion matrix SVM RBF & Skip-Gram.....	70
Gambar 4.29. Confusion matrix SVM Sigmoid & CBOW	72
Gambar 4.30. confusion matrix SVM Sigmoid & Skip-Gram	74
Gambar 4.31. ROC Curve SVM RBF & Skip-Gram (SMOTE)	76
Gambar 4.32. ROC Curve SVM Sigmoid & CBOW	77
Gambar 4.33. SVM Sigmoid & Skip-Gram.....	77
Gambar 4.34. Tampilan Fitur Download Template dan File Reader	78
Gambar 4.35. Tampilan Fitur Tabel Keseluruhan Data.....	79
Gambar 4.36. Tampilan Fitur Tabel Hasil Analisis Data	79

DAFTAR TABEL

Tabel 2.1. Confusion Matrix	18
Tabel 2.2. Tingkatan Nilai AUC.....	20
Tabel 3.1. Sample Data	28
Tabel 3.2. Perbedaan Prediksi CBOW & Skip-Gram	32
Tabel 4.1 Data hasil scraping	38
Tabel 4.2. Hasil pelabelan data	43
Tabel 4.3. Hasil cleansing data.....	44
Tabel 4.4. Hasil case folding.....	46
Tabel 4.5. Hasil stopword removal	48
Tabel 4.6. Hasil tokenizing data.....	49
Tabel 4.7. Hasil stemming data	51
Tabel 4.8. Classification report SVM Linear & CBOW (SMOTE).....	59
Tabel 4.9. Classification report SVM Linear & CBOW (Tanpa SMOTE)	59
Tabel 4.10. Classification report SVM Linear & Skip-Gram (SMOTE)	61
Tabel 4.11. Classification report SVM Linear & Skip-Gram (Tanpa SMOTE)....	61
Tabel 4.12. Classification report SVM polynomial & CBOW (SMOTE)	63
Tabel 4. 13. Classification report SVM polynomial & CBOW (Tanpa SMOTE)	63
Tabel 4.14. Classification report SVM polynomial & Skip-Gram (SMOTE)	65
Tabel 4.15. Classification report SVM polynomial & Skip-Gram (Tanpa SMOTE)	65
Tabel 4.16. Classification report SVM RBF & CBOW (SMOTE).....	67
Tabel 4.17. Classification report SVM RBF & CBOW (Tanpa SMOTE)	67
Tabel 4.18. Classification report SVM RBF & Skip-Gram (SMOTE).....	69
Tabel 4.19. Classification report SVM RBF & Skip-Gram (Tanpa SMOTE)	69
Tabel 4.20. Classification report SVM sigmoid & CBOW (SMOTE)	71
Tabel 4.21. Classification report SVM sigmoid & CBOW (Tanpa SMOTE).....	71
Tabel 4.22. Classification report SVM sigmoid & Skip-Gram (SMOTE).....	73
Tabel 4.23. Classification report SVM sigmoid & Skip-Gram (Tanpa SMOTE). 73	
Tabel 4.24. Hasil klasifikasi secara keseluruhan.....	75
Tabel 4.25. Confusion Matrix Data Validasi	80
Tabel 4.26. Perbandingan Sentimen Hasil Prediksi Sistem	80