

**SKRIPSI**



**ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI MAXIM  
MENGUNAKAN METODE KLASIFIKASI *RANDOM FOREST* DAN  
EKSTRAKSI FITUR *WORD2VEC***

**Rafid Ammar Adinegoro  
NIM. 1910511076**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
2023**

**SKRIPSI**



**ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI MAXIM  
MENGUNAKAN METODE KLASIFIKASI *RANDOM FOREST* DAN  
EKSTRAKSI FITUR *WORD2VEC***

**Rafid Ammar Adinegoro  
NIM. 1910511076**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
2023**

## PERNYATAAN ORISINALITAS

### PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Rafid Ammar Adinegoro

NRP : 191511076

Tanggal : 19 Desember 2023

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 19 Desember 2023

Yang Menyatakan



(Rafid Ammar Adinegoro)

## **PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

### **PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademis Universitas Pembangunan Nasional Veteran Jakarta, Saya bertanda tangan dibawah ini :

Nama : Rafid Ammar Adinegoro

NRP : 191511076

Fakultas : Ilmu Komputer

Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

#### **ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI MAXIM MENGUNAKAN METODE KLASIFIKASI *RANDOM FOREST* DAN EKSTRAKSI FITUR *WORD2VEC***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 19 Desember 2023

Yang Menyatakan,



(Rafid Ammar Adinegoro)

## LEMBAR PENGESAHAN

### LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

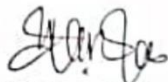
Nama : Rafid Ammar Adinegoro

NIM : 191511076

Program Studi : S1 Informatika

Judul Skripsi/TA : Analisis Sentimen Pada Ulasan Pengguna Aplikasi Maxim Menggunakan Metode Klasifikasi *Random Forest* Dan Ekstraksi Fitur *Word2Vec*.

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.



Yuni Widiastiwati, S.Kom, M.Si.,

Penguji 1



Zatin Niqotaini, S.Tr.Kom., M.Kom

Penguji 2



Henki Bayu Seta, S.Kom., MTL

Pembimbing 1



Iin Ernawati, S.Kom., M.Si.

Pembimbing 2



Erly Krisnanik, S.Kom.,MM.

Pt. Dekan



Dr. Widya Cholil, M.I.T.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 30 November 2023



# ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI MAXIM MENGUNAKAN METODE KLASIFIKASI *RANDOM FOREST* DAN EKSTRAKSI FITUR *WORD2VEC*

Rafid Ammar Adinegoro

## ABSTRAK

Revolusi industri 4.0 menyebabkan bisnis mulai bergerak ke ranah digital, salah satunya adalah transportasi online. Beberapa perusahaan transportasi online mulai bermunculan dengan meluncurkan bisnisnya melalui aplikasi, salah satunya adalah Maxim. Seiring dengan peningkatan jumlah pengguna aplikasi, maka tuntutan terhadap kualitas layanan aplikasi juga meningkat. Salah satu cara untuk menilai suatu layanan aplikasi adalah dengan melihat ulasan pengguna. Pada penelitian ini dilakukan analisis sentimen terhadap 2 kelas positif dan negatif pada ulasan pengguna aplikasi Maxim. Algoritma klasifikasi yang digunakan adalah *Random Forest* serta menerapkan ekstraksi fitur *Word2vec*. Ekstraksi fitur *Word2vec* digunakan agar analisis sentimen menjadi lebih efektif karena mampu mengenali semantik antar kata. Dengan menguji nilai parameter *Word2vec* yang berbeda, didapatkan performa terbaik pada parameter dimensi 300, *window* 7, dan *epochs* 10. Hasil dari penelitian ini membuktikan bahwa klasifikasi *Random Forest* dengan ekstraksi fitur *Word2vec* mendapatkan tingkat performa lebih besar dibandingkan tanpa menggunakan ekstraksi fitur *Word2vec* dimana menghasilkan nilai akurasi sebesar 93,39%, presisi sebesar 95,85%, *recall* sebesar 91,54%, dan *f1-score* sebesar 93,65%. Dibanding dengan tingkat performa tanpa menggunakan ekstraksi fitur hanya menghasilkan nilai akurasi sebesar 92,63%, presisi sebesar 95,79%, *recall* sebesar 90,10% dan *f1-score* sebesar 92,86%.

**Kata Kunci:** Analisis Sentimen, Maxim, Klasifikasi, *Random Forest*, *Word2Vec*

# SENTIMENT ANALYSIS ON MAXIM APPLICATION USER REVIEWS USING RANDOM FOREST CLASSIFICATION METHOD AND WORD2VEC FEATURE EXTRACTION

Rafid Ammar Adinegoro

## ABSTRACT

The industrial revolution 4.0 has caused businesses to start moving into the digital realm, one of which is online transportation. Several online transportation companies have started to appear by launching their business through applications, one of which is Maxim. Along with the increase in the number of application users, the demand for application service quality has also increased. One way to assess an application service is to look at user reviews. This research conducted sentiment analysis on 2 classes, positive and negative, in user reviews of the Maxim application. The classification algorithm used is Random Forest and applies Word2vec feature extraction. Word2vec feature extraction is used to make sentiment analysis more effective because it is able to recognize semantics between words. By testing different Word2vec parameter values, the best performance was obtained at dimension parameters 300, window 7, and epochs 10. The results of this study demonstrate that the *Random Forest* classification with *Word2Vec* feature extraction achieved a higher level of performance compared to not using *Word2Vec* feature extraction, resulting in an accuracy of 93,39%, precision of 95,85%, recall of 91,54%, and an F1-score of 93,65%. In comparison, the performance without using feature extraction only yielded an accuracy of 92,63%, precision of 95,79%, recall of 90,10%, and an F1-score of 92,86%.

**Keyword:** *Sentiment Analysis, Maxim, Classification, Random Forest, Word2Vec*

## KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat dan karunia-nya sehingga penulis dapat menyelesaikan Skripsi. Skripsi ini disusun sebagai syarat dari Skripsi/Tugas Akhir Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jakarta.

Dengan ini penulis menyadari bahwa banyak pihak yang memberi dukungan dan bantuan selama pengerjaan Skripsi ini, oleh karena itu dengan penuh rasa hormat Penulis ingin mengucapkan terima kasih kepada:

1. Bapak Henki Bayu Seta, S.Kom., MTI. selaku dosen pembimbing 1.
2. Ibu Iin Ernawati, S.Kom., M.Si. selaku dosen pembimbing 2.
3. Ibu Erly Krisnanik, S.Kom., MM selaku Plt. Dekan Fakultas Ilmu Komputer.
4. Ibu Dr. Widya Cholil, M.I.T. selaku Ketua Program Studi Sarjana Jurusan S1 Informatika.
5. Seluruh Bapak/Ibu Dosen Pengajar Fakultas Ilmu Komputer atas ilmu-ilmu bermanfaat yang diberikan.
6. Kedua orang tua yang telah memberikan dukungan baik secara moril maupun materil.
7. Teman-teman, saudara dan pihak-pihak lain yang tidak dapat penulis sebutkan satu per satu, terima kasih atas saran dan dukungan yang telah diberikan.

Jakarta, 18 Oktober 2023

Penulis



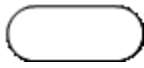


## DAFTAR ISI

LEMBAR JUDUL.....	i
PERNYATAAN ORISINALITAS.....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS.....	iv
LEMBAR PENGESAHAN.....	v
ABSTRAK.....	vi
ABSTRACT.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR SIMBOL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN.....	xvi
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
<b>1.1. Latar Belakang.....</b>	<b>1</b>
<b>1.2. Rumusan Masalah.....</b>	<b>3</b>
<b>1.3. Ruang Lingkup.....</b>	<b>3</b>
<b>1.4. Tujuan Penelitian.....</b>	<b>3</b>
<b>1.5. Luaran Yang Diharapkan.....</b>	<b>4</b>
<b>1.6. Manfaat Penelitian.....</b>	<b>4</b>
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>5</b>
<b>2.1. Text Mining.....</b>	<b>5</b>
<b>2.2. Text Preprocessing.....</b>	<b>5</b>
<b>2.2.1. Case Folding.....</b>	<b>5</b>
<b>2.2.2. Data Cleaning.....</b>	<b>5</b>
<b>2.2.3. Stemming.....</b>	<b>6</b>
<b>2.2.4. Tokenizing.....</b>	<b>6</b>
<b>2.2.5. Stopword Removal.....</b>	<b>6</b>
<b>2.2.6. Normalization.....</b>	<b>6</b>
<b>2.3. Analisis Sentimen.....</b>	<b>6</b>
<b>2.4. Web Scraping.....</b>	<b>7</b>
<b>2.5. Pembobotan TF IDF.....</b>	<b>7</b>
<b>2.6. Algoritma Random Forest.....</b>	<b>8</b>
<b>2.7. Ekstraksi Fitur Word2Vec.....</b>	<b>10</b>
<b>2.8. Confusion Matrix.....</b>	<b>12</b>

2.9.	ROC Curve .....	14
2.10.	Maxim.....	14
2.11.	Review Penelitian Terdahulu .....	15
<b>BAB III METODE PENELITIAN .....</b>		<b>18</b>
3.1.	Kerangka Pikir .....	18
3.1.1.	Identifikasi Masalah.....	19
3.1.2.	Studi Literatur.....	19
3.1.3.	Akuisisi Data.....	19
3.1.4.	Pelabelan Data.....	20
3.1.5.	Cohen's Kappa <i>Statistics</i> .....	20
3.1.6.	<i>Text Preprocessing</i> .....	21
3.1.6.1	<i>Case Folding</i> .....	21
3.1.6.2	<i>Data Cleaning</i> .....	22
3.1.6.3	<i>Stemming</i> .....	22
3.1.6.4	<i>Tokenizing</i> .....	22
3.1.6.5	<i>Stopword Removal</i> .....	22
3.1.6.6	<i>Normalization</i> .....	22
3.1.7.	Pembobotan TF-IDF .....	22
3.1.8.	Ekstraksi Fitur <i>Word2Vec</i> .....	23
3.1.9.	Klasifikasi Random Forest .....	24
3.1.10.	Evaluasi Model .....	24
3.1.11.	Visualisasi.....	25
3.2.	Perangkat Penelitian .....	25
3.3.	Jadwal Pelaksanaan Penelitian .....	26
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>27</b>
4.1	Pengumpulan Data .....	27
4.2	<i>Labeling</i> .....	28
4.3	<i>Text Pre-Processing</i> .....	32
4.3.1.	<i>Case Folding</i> .....	32
4.3.2	<i>Data Cleaning</i> .....	33
4.3.3	<i>Stemming</i> .....	34
4.3.4	<i>Normalization</i> .....	35
4.3.5	<i>Stopword Removal</i> .....	35
4.3.6	<i>Tokenizing</i> .....	36
4.4	Pembagian Data .....	38
4.5	Ekstraksi Fitur <i>Word2Vec</i> .....	39

<b>4.6</b>	<b>Pembobotan TF-IDF .....</b>	<b>43</b>
<b>4.7</b>	<b>Klasifikasi <i>Random Forest</i> .....</b>	<b>47</b>
<b>4.8</b>	<b>Evaluasi .....</b>	<b>49</b>
	<b>4.8.1.1. Evaluasi Parameter Dimensi <i>Word2vec</i> Pada Model Klasifikasi .....</b>	<b>50</b>
	<b>4.8.1.2. Evaluasi Parameter <i>Window Word2vec</i> Pada Model Klasifikasi .....</b>	<b>57</b>
	<b>4.8.1.3. Evaluasi Parameter <i>Epochs Word2vec</i> Pada Model Klasifikasi .....</b>	<b>64</b>
	<b>4.8.2. Evaluasi Klasifikasi Model TF-IDF .....</b>	<b>77</b>
	<b>4.8.3. Perbandingan Model <i>Word2Vec</i> dan TF-IDF .....</b>	<b>80</b>
<b>4.9</b>	<b>Visualisasi .....</b>	<b>80</b>
	<b>4.9.1. Visualisasi Sentimen Positif .....</b>	<b>81</b>
	<b>4.9.2. Visualisasi Sentimen Negatif .....</b>	<b>82</b>
<b>BAB V</b>	<b>PENUTUP .....</b>	<b>84</b>
	<b>5.1. Kesimpulan .....</b>	<b>84</b>
	<b>5.2. Saran .....</b>	<b>84</b>
	<b>DAFTAR PUSTAKA .....</b>	<b>86</b>
	<b>LAMPIRAN .....</b>	<b>89</b>
	<b>RIWAYAT HIDUP .....</b>	<b>119</b>

## DAFTAR SIMBOL

Simbol Flowchart			
NO	GAMBAR	NAMA	KETERANGAN
1		<i>Terminal</i> ( <i>start,</i> <i>end</i> )	Menggambarkan bagaimana kegiatan dimulai atau kegiatan berakhir.
2		<i>Flow</i> <i>Direction</i>	Menggambarkan hubungan antar simbol yang menyatakan suatu jalannya proses dalam sistem.
3		<i>Process</i>	Menggambarkan deskripsi dari proses yang dijalankan.

## DAFTAR GAMBAR

Gambar 2.1 Cara Kerja Algoritma <i>Random Forest</i> (Aufa & Qoiriah, 2022) .....	9
Gambar 2.2 Cara Kerja Pembobotan Kata <i>Word2Vec</i> (Aufa & Qoiriah, 2022)....	11
Gambar 3.1. Kerangka Pikir.....	18
Gambar 3.2. Interpretasi Kappa (Qureshi et al., 2022) .....	21
Gambar 4. 1 Data Hasil Web Scraping .....	27
Gambar 4. 2 Persentase Kelas Positif dan Negatif.....	32
Gambar 4. 3 Tokenizing Sebelum Normalization.....	37
Gambar 4. 4 Tokenizing Setelah Normalization dan Stopword Removal .....	38
Gambar 4. 5 Ilustrasi Window Model Skip-gram.....	40
Gambar 4. 6 Confusion Matrix dengan Word2vec 100_5_10 .....	50
Gambar 4. 7 Confusion Matrix dengan Word2vec 100_7_15 .....	52
Gambar 4. 8 Confusion Matrix dengan Word2vec 300_5_10 .....	54
Gambar 4. 9 Confusion Matrix dengan Word2vec 300_7_15 .....	55
Gambar 4. 10 Perbandingan Rata-Rata Evaluasi Model Berdasarkan Parameter Dimensi.....	57
Gambar 4. 11 Confusion Matrix dengan Word2vec 100_5_35 .....	58
Gambar 4. 12 Confusion Matrix dengan Word2vec 100_7_35 .....	60
Gambar 4. 13 Confusion Matrix dengan Word2vec 100_11_35 .....	62
Gambar 4. 14 Perbandingan Rata-Rata Evaluasi Model Berdasarkan Parameter Window.....	64
Gambar 4. 15 Confusion Matrix dengan Word2vec 300_11_10 .....	65
Gambar 4. 16 Confusion Matrix dengan Word2vec 300_11_15 .....	67
Gambar 4. 17 Confusion Matrix dengan Word2vec 300_11_35 .....	69
Gambar 4. 18 Rata-Rata Performa Model Berdasarkan Parameter Epochs.....	71
Gambar 4. 19 Confusion Matrix dengan Word2vec 300_11_5 .....	72
Gambar 4. 20 Perbandingan Performa Model Berdasarkan Parameter Epochs Setelah Penambahan.....	74
Gambar 4. 21 Confusion Matrix dengan Word2vec 300_7_10 .....	75
Gambar 4. 22 Grafik ROC Curve Model Random Forest dengan Word2vec .....	77
Gambar 4. 23 Confusion Matrix dengan TF-IDF .....	78
Gambar 4. 24 Grafik ROC Curve Model Random Forest dengan TF-IDF .....	79

Gambar 4. 25 Visualisasi Sentimen Positif.....	81
Gambar 4. 26 Frekuensi Kata Sentimen Positif.....	82
Gambar 4. 27 Visualisasi Sentimen Negatif .....	82
Gambar 4. 28 Frekuensi Kata Sentimen Negatif .....	83

## DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i> .....	13
Tabel 3.1 Jadwal Pelaksanaan Penelitian.....	26
Tabel 4. 1 Sampel Pelabelan Ulasan .....	28
Tabel 4. 2 Hasil Pelabelan Data Ulasan .....	29
Tabel 4. 3 Contoh Ulasan Sebelum dan Sesudah Case Folding .....	33
Tabel 4. 4 Contoh Ulasan Sebelum dan Sesudah Data Cleaning.....	34
Tabel 4. 5 Contoh Ulasan Sebelum dan Sesudah Stemming .....	34
Tabel 4. 6 Contoh Ulasan Sebelum dan Sesudah Normalization.....	35
Tabel 4. 7 Contoh Ulasan Sebelum dan Sesudah Stopword Removal.....	36
Tabel 4. 8 Contoh Ulasan Sebelum dan Sesudah Tokenizing .....	36
Tabel 4. 9 Pembagian Data Latih dan Data Uji .....	39
Tabel 4. 10 Bentuk Vektor Biner dengan Metode one-hot-encoder .....	39
Tabel 4. 11 Hasil Ekstraksi Fitur Word2vec .....	42
Tabel 4. 12 Vektor Kalimat.....	42
Tabel 4. 13 Nilai Parameter Word2vec.....	43
Tabel 4. 14 Sampel Ulasan Penerapan TF-IDF .....	44
Tabel 4. 15 Perhitungan Term Frequency .....	44
Tabel 4. 16 Perhitungan Inverse Document Frequency .....	45
Tabel 4. 17 Perhitungan Pembobotan TF-IDF .....	46
Tabel 4. 18 Model Dataset .....	47
Tabel 4. 19 Hasil Evaluasi Model Berdasarkan Parameter Dimensi 100 .....	53
Tabel 4. 20 Hasil Evaluasi Model Berdasarkan Parameter Dimensi 300 .....	56
Tabel 4. 21 Hasil Evaluasi Model Berdasarkan Parameter Window 5 .....	59
Tabel 4. 22 Hasil Evaluasi Model Berdasarkan Parameter Window 7 .....	61
Tabel 4. 23 Hasil Evaluasi Model Berdasarkan Parameter Window 11 .....	63
Tabel 4. 24 Hasil Evaluasi Model Berdasarkan Parameter Epochs 10 .....	66
Tabel 4. 25 Hasil Evaluasi Model Berdasarkan Parameter Epochs 15 .....	68
Tabel 4. 26 Hasil Evaluasi Model Berdasarkan Parameter Epochs 35 .....	70
Tabel 4. 27 Hasil Evaluasi Model Berdasarkan Parameter Epochs 5 .....	73
Tabel 4. 28 Hasil Performa Model Klasifikasi Menggunakan Word2vec .....	76
Tabel 4. 29 Perbandingan Performa Model Word2vec dengan TF-IDF.....	80

## DAFTAR LAMPIRAN

Nomor	Judul
1.	Korpus Normalization
2.	Korpus Stopword Tambahan
3.	Source Code
4.	Daftar Rumus
5.	Hasil Plagiarisma