

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN MAXIMUM
ENTROPY UNTUK ANALISIS SENTIMEN PADA ULASAN APLIKASI MY
PERTAMINA DI GOOGLE PLAYSTORE MENGGUNAKAN PYTHON
SKRIPSI**



Oleh :

RAHMAT AULIYA

18081010027

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

2023

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN MAXIMUM
ENTROPY UNTUK ANALISIS SENTIMEN PADA ULASAN APLIKASI MY
PERTAMINA DI GOOGLE PLAYSTORE MENGGUNAKAN PYTHON**

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh

Gelar Sarjana Komputer Program Studi Informatika



Oleh :

RAHMAT AULIYA

18081010027

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2023

LEMBAR PENGESAHAN

**Judul : PERBANDINGAN ALGORITMA RANDOM FOREST DAN
MAXIMUM ENTROPY UNTUK ANALISIS SENTIMEN PADA
ULASAN APLIKASI MY PERTAMINA DI GOOGLE
PLAYSTORE MENGGUNAKAN PYTHON**

Oleh : RAHMAT AULIYA

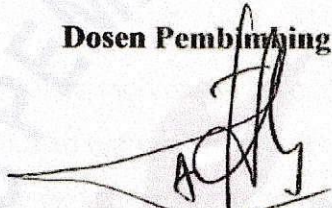
NPM : 18081010027

**Telah Diseminarkan Dalam Ujian Skripsi Pada :
Hari Rabu Tanggal 12 Juli 2023**

Mengetahui

Dosen Pembimbing

1.

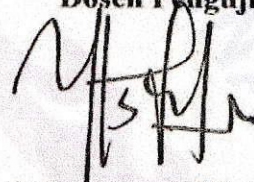


Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

Dosen Penguji

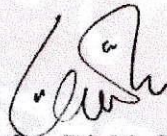
1.



Yisti Vita Via, S.ST, M.Kom

NIP. 19860425 2021212 001

2.



Agung Mustika Rizki, S.Kom, M.Kom

NIP. 19930725 202203 1008

2.



Faywaz Ali Akbar, S.Kom, M.Kom

NIP. 19920317 2018031 002

Menyetujui

Dekan

Fakultas Ilmu Komputer

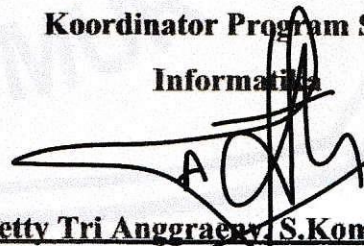


Dr. Novirina Hendrasarie, S.T, M.T

NIP. 19681126 199403 2 001

Koordinator Program Studi

Informatika



Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa program studi Informatika UPN “Veteran” Jawa Timur, yang bertanda tangan dibawah ini :

Nama : Rahmat Auliya

NPM : 18081010027

Menyatakan bahwa judul skripsi yang saya ajukan dan kerjakan yang berjudul :

“PERBANDINGAN ALGORITMA RANDOM FOREST DAN MAXIMUM ENTROPY UNTUK ANALISIS SENTIMEN PADA ULASAN APLIKASI MY PERTAMINA DI GOOGLE PLAYSTORE MENGGUNAKAN PYTHON

”

Bukan merupakan plagiat dari skripsi/tugas akhir/penelitian orang lain dan juga bukan merupakan produk atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi pendidikan lainnya.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima segala konsekuensinya.

Surabaya, 4 Juli 2023



RAHMAT AULIYA

NPM. 18081010027

PERBANDINGAN ALGORITMA RANDOM FOREST DAN MAXIMUM ENTROPY UNTUK ANALISIS SENTIMEN PADA ULASAN APLIKASI MY PERTAMINA DI GOOGLE PLAYSTORE MENGGUNAKAN PYTHON

Nama Mahasiswa : Rahmat Auliya

NPM : 18081010027

Program Studi : Informatika

Dosen Pembimbing : Fetty Tri Anggraeny, S.Kom, M.Kom

Agung Mustika Rizki, S.Kom, M.Kom

ABSTRAK

Dalam dunia digital yang terus berkembang, ulasan pengguna pada platform seperti Google PlayStore memiliki peran penting dalam membentuk persepsi dan keputusan pengguna terkait dengan sebuah aplikasi. Untuk memahami sentimen dan pandangan pengguna terhadap aplikasi My Pertamina, analisis sentimen pada ulasan pengguna menjadi penting. Maka dari itu penelitian ini bertujuan untuk mengetahui gambaran umum data ulasan tentang aplikasi My Pertamina berdasarkan situs Google Play dan seberapa tepat machine learning dengan menggunakan metode Random Forest dan Maximum Entropy dalam mengklasifikasikan teks berbahasa Indonesia mengenai ulasan pengguna aplikasi My Pertamina berdasarkan situs Google Play. Metode yang digunakan dalam penelitian ini adalah Random Forest dan Maximum Entropy. Penelitian ini dilakukan dengan menggunakan data ulasan aplikasi My Pertamina di Google Playstore sebanyak 1.000 data ulasan. Data yang dimiliki dibagi menjadi sembilan skenario perbandingan data train dan tes. Percobaan pada Random Forest memiliki akurasi lebih unggul dengan nilai 99,62%, 99,62%, 99,25%, 99,16%, 98,65%, 98,75, 97,06%, 94,26%, dan 88,26%. Sedangkan untuk Maxium Entropy memiliki nilai akurasi 98,88%, 98,50%, 98,50%, 98,69%, 97,76%, 96,57%, 95,30%, 94,16%, 89,46%.

Kata kunci: *Sentimen analisis, Random Forest, Maximum Entropy, My Pertamina, Google Playstore*

KATA PENGANTAR

Dengan mengucapkan puji syukur kehadiran Allah SWT yang telah memberikan petunjuk, pertolongan serta seluruh nikmat-Nya kepada kita, dengan izin dan ridho-Nya lah penulis dapat menyelesaikan laporan skripsi ini yang berjudul “Perbandingan Algoritma Random Forest Dan Maximum Entropy Untuk Analisis Sentimen Pada Ulasan Aplikasi My Pertamina Di Google Playstore Menggunakan Python”

Banyak dukungan dan bantuan yang didapatkan selama melakukan penelitian hingga akhirnya mampu menyelesaikan penulisan laporan skripsi ini. Dengan rasa hormat, penulis mengucapkan terima kasih kepada seluruh pihak terkait yang turut membantu dan terlibat dalam penyusunan laporan ini dari awal hingga akhir.

Penulis menyadari mungkin masih ada kekurangan pada laporan skripsi ini mengingat keterbatasan pengetahuan dan kemampuan. Oleh karena itu, penulis sangat mengharapkan masukan, kritik, dan saran yang dapat membangun untuk mengembangkan laporan ini.

Surabaya, 4 Juli 2023

Penyusun

UCAPAN TERIMA KASIH

Dalam penyusunan dan penulisan skripsi ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari beberapa pihak. Oleh karena itu dalam kesempatan ini penulis menyampaikan terima kasih kepada Allah SWT yang selalu ada dan memberi kelancaran, kemudahan, kekuatan dan sumber dari segala pencerahan.

Tidak lupa penulis berterima kasih kepada yang terhormat :

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Dr. Novirina Hendrasarie, S.T, M.T. Selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Fetty Tri Anggraeny, S.Kom, M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Fetty Tri Anggraeny, S.Kom, M.Kom. selaku dosen pembimbing I dan Bapak Agung Mustika Rizki, S.Kom., M.Kom.. selaku dosen pembimbing II, yang telah sabar, ikhlas meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, motivasi, arahan, dan saran-saran yang sangat berharga kepada penulis selama penyusunan skripsi ini.
5. Staff dan dosen Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmu, arahan, dan pengalamannya selama perkuliahan.
6. Kedua orang tua penulis yang selalu memberi semangat, motivasi, dan mendoakan untuk kebaikan, keberhasilan maupun kesuksesan penulis.
7. Teman-teman Teknik Informatika angkatan 2018, yang sudah bersedia membantu, mengisi, dan menjadi bagian dalam cerita hidup dan perkuliahan penulis.
8. Semua pihak yang tidak dapat penulis ucapkan yang juga turut mendoakan dan menyemangati penulis dalam menyelesaikan skripsi ini.

Terima kasih atas segala bantuan, dukungan dan doa-doanya semoga Allah SWT memberikan balasan yang setimpal kepada semua pihak yang telah membantu penulis sehingga penulis mampu dan berhasil menyelesaikan penelitian dan proses penyusunan laporan penelitian skripsi dengan baik dan tepat.

Surabaya, 4 Juli 2023

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
ABSTRAK	iii
KATA PENGANTAR.....	iv
UCAPAN TERIMA KASIH	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu	5
2.2 Analisis Sentimen	6
2.3 My Pertamina.....	7
2.4 Google Play.....	7
2.5 Data Mining	8
2.6 Text Mining	9
2.7 Text Preprocessing.....	10
2.7.1 Case Folding	10
2.7.2 Stopword Removal.....	11
2.7.3 Tokenizing	12
2.7.4 Stemming	12
2.8 WordCloud.....	13
2.9 Ekstraksi Fitur.....	13

2.9.1	TF-IDF	14
2.10	Random Forest	15
2.11	Maximum Entropy	17
2.12	Confusion Matrix	18
2.12.1	Accuracy	20
2.12.2	Precision.....	20
2.12.3	Recal	20
2.12.4	F1-Score.....	21
2.13	Tuning Parameter.....	21
2.13.1	GridSearchCV.....	21
BAB III METODOLOGI.....		23_Toc141091264
3.1	Tahapan pengerjaan tugas akhir.....	23
3.2	Studi Literatur	23
3.3	Analisa dan Desain	24
3.4	Scraping Data.....	25
3.5	Analisis Deskriptif dan Lebellng.....	25
3.6	Preprocessing dan Ekstraksi Fitur.....	28
3.7	Klasifikasi Data.....	37
3.7.1	Klasifikasi Random Forest.....	37
3.7.2	Klasifikasi Maximum Entropy.....	39
3.8	Pengujian Sistem dengan Confusion Matrix.....	40
BAB IV HASIL DAN PEMBAHASAN		41
4.1	Lingkungan Penelitian	41
4.1.1	Spesifikasi Perangkat keras.....	41
4.1.2	Spesifikasi Perangkat Lunak.....	41
4.2	Proses pengambilan data.....	42
4.3	Data Framing	43
4.4	Pelebelan Kelas Sentimen.....	47
4.5	Proccessing Data.....	51
4.5.1	Case Folding	51

4.5.2	Stopword Removal.....	52
4.5.3	Tokenizing	53
4.5.4	Steaming	54
4.6	WordCloud.....	56
4.6.1	WordCloud Positif	56
4.6.2	WordCloud Negatif.....	57
4.6.3	WordCloud Netral.....	58
4.7	Import Library.....	59
4.8	TF-IDF	60
4.9	Handling Imbalance	61
4.10	Split train dan test	61
4.11	Proses Klasifikasi.....	63
4.12	Proses Pengujian	64
4.13	Tuning Parameter.....	73
BAB V PENUTUP.....		76
5.1	Kesimpulan	76
5.2	Saran	76
DAFTAR PUSTAKA.....		78

DAFTAR TABEL

Tabel 2.1 Case Folding	11
Tabel 2.2 Stopword Removal.....	11
Tabel 2.3 Tokenizing	12
Tabel 2.4 Stemming	13
Tabel 2.5 Multiclass Confusion Matrix	19
Tabel 3.1 Tahapan Pengerjaan Tugas Akhir	23
Tabel 3.2 Rating pengguna terhadap e-commerce My Pertamina pada situs Google Play.....	26
Tabel 3.3 Jumlah ulasan positif, netral dan negatif terhadap e-commerce My Pertamina.....	27
Tabel 3.4 Hasil Flowchart alur proses Case Folding.....	30
Tabel 3.5 Hasil Flowchart alur proses Stopword Removal	31
Tabel 3.6 Hasil Flowchart alur proses Tokenizing.....	33
Tabel 3.7 Hasil Flowchart alur proses Stemming	34
Tabel 3.8 Dokumen Term	36
Tabel 3.9 Indeks Kata	40
Tabel 4.1 Spesifikasi perangkat keras yang digunakan	41
Tabel 4.2 Spesifikasi perangkat lunak yang digunakan	42
Tabel 4.3 Split train dan test	61
Tabel 4.4 Hasil precision, recall, dan f1-score.....	69
Tabel 4.5 Perbandingan Akurasi.....	70
Tabel 4.6 Confusion Matrix	73

DAFTAR GAMBAR

Gambar 2.1 Tahapan Text Preprocessing	10
Gambar 2.2 Algoritma Random Forest.....	16
Gambar 3.1 Flowchart Penelitian.....	24
Gambar 3.2 Rating pengguna terhadap e-commerce My Pertamina pada situs Google Play.....	24
Gambar 3.3 Jumlah ulasan positif, netral dan negatif terhadap e-commerce My Pertamina.....	24
Gambar 3.4 Pemisahan data text untuk tahap preprocessing	24
Gambar 3.5 Tahap Preprocessing	24
Gambar 3.6 Ulasan My Pertamina Sebelum Processing	24
Gambar 3.7 Flowchart alur proses Case Folding.....	24
Gambar 3.8 Flowchart alur proses Stopward Removal	24
Gambar 3.9 Flowchart alur proses Tokenizing.....	24
Gambar 3.10 Flowchart alur proses Stemming	24
Gambar 3.11 File csv hasil preprocessing.....	24
Gambar 3.12 Tahapan Ekstraksi fitur TF-IDF.....	24
Gambar 3.13 Tahapan klasifikasi.....	24
Gambar 4.1 Hasil Scraping Data Ulasan Aplikasi My Pertamina	43
Gambar 4.2 Lima data teratas shorteddata.....	44
Gambar 4.3 Menambahkan Kolom Year, Mount dan Day	45
Gambar 4.4 Dataframe df.....	45
Gambar 4.5 Rekapitulasi score	46
Gambar 4.6 Distribusi Score.....	47
Gambar 4.7 Menambahkan Kolom Sentimen.....	48
Gambar 4.8 Distribusi Sentimen	50
Gambar 4.9 Pie Chart Distribusi Sentimen.....	75
Gambar 4.10 Distribusi Sentiemen Per Tahun.....	50
Gambar 4.11 Distribusi Sentiemen Per Tahun.....	50
Gambar 4.12 Menambahkan Kolom text_clean Hasil Case Folding.....	52

Gambar 4.13 Menambahkan Kolom text_stopword Hasil Stopword Removal.....	53
Gambar 4.14 Menambahkan Kolom text_tokens Hasil Tokenizing.....	54
Gambar 4.15 Data setelah melalui proses Stemming	55
Gambar 4.16 Menambahkan Kolom text_steamindo Hasil Steaming	56
Gambar 4.17 Wordcloud Sentiment Positif	57
Gambar 4.18 Wordcloud Sentiment Negatif.....	58
Gambar 4.19 WordCloud Sentiment Netral.....	59
Gambar 4.20 Random Forest train 90% dan test 10%	64
Gambar 4.21 Maximum Entropy train 90% dan test 10%	64
Gambar 4.22 Random Forest train 80% dan test 20%	65
Gambar 4.23 Maximum Entropy train 80% dan test 20%	65
Gambar 4.24 Random Forest train 70% dan test 30%	65
Gambar 4.25 Maximum Entropy train 70% dan test 30%	65
Gambar 4.26 Random Forest train 60% dan test 40%	66
Gambar 4.27 Maximum Entropy train 60% dan test 40%	66
Gambar 4.28 Random Forest train 50% dan test 50 %	66
Gambar 4.29 Maximum Entropy train 50% dan test 50%	66
Gambar 4.30 Random Forest train 40% dan test 60%	67
Gambar 4.31 Maximum Entropy train 40% dan test 60%	67
Gambar 4.32 Random Forest train 30% dan test 70%	67
Gambar 4.33 Maximum Entropy train 30% dan test 70%	67
Gambar 4.34 Random Forest train 20% dan test 80%	68
Gambar 4.35 Maximum Entropy train 20% dan test 80%	68
Gambar 4.36 Random Forest train 10% dan test 90%	68
Gambar 4.37 Maximum Entropy train 10% dan test 90%	68
Gambar 4.38 Plot Line	71
Gambar 4.39 Plot Bar.....	72
Gambar 4.40 Tuning Parameter Random Forest.....	75
Gambar 4.41 Tuning Parameter Maximum Entropy.....	75