

**PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI
UNTUK ANALISIS SENTIMEN KOMENTAR SHOPEE DI
GOOGLE PLAY STORE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

KHUSWATUN HASANAH

20.11.3667

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI
UNTUK ANALISIS SENTIMEN KOMENTAR SHOPEE DI
GOOGLE PLAY STORE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

KHUSWATUN HASANAH

20.11.3667

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI
UNTUK ANALISIS SENTIMEN KOMENTAR SHOPEE DI
GOOGLE PLAY STORE**

yang disusun dan diajukan oleh

Khuswatun Hasanah

20.11.3667

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Januari 2024

Dosen Pembimbing,



Ainul Yaqin, M.KOM

NIK. 190302255

HALAMAN PENGESAHAN
SKRIPSI
PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI
UNTUK ANALISIS SENTIMEN KOMENTAR SHOPEE DI
GOOGLE PLAY STORE

yang disusun dan diajukan oleh

KHUSWATUN HASANAH

20.11.3667

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Januari 2024

Susunan Dewan Penguji

Nama Penguji

Nuri Cahyono, M.Kom
NIK. 190302278

Anggit Ferdita Nugraha, S.T., M.Eng.
NIK. 190302480

Ainul Yaqin, M. Kom
NIK. 190302255

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Januari 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Khuswatun Hasanah
NIM : 20.11.3667

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI UNTUK ANALISIS SENTIMEN KOMENTAR SHOPEE DI GOOGLE PLAY STORE

Dosen Pembimbing : Ainul Yaqin, M. Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Januari 2024

Yang Menyatakan,



METERAI TEMPEL
E4CALX042812236

Khuswatun Hasanah

HALAMAN PERSEMBAHAN

Skripsi ini disusun dan dipersembahkan dengan penuh rasa syukur dan cinta kepada :

1. Cinta pertama dan panutanku, Bapak Eko Saputra. Beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan, namun beliau mampu mendidik saya, memotivasi, memberikan dukungan hingga saya mampu menyelesaikan study sampai sarjana.
2. Pintu surgaku, Ibu Surati. Beliau sangat berperan penting dalam menyelesaikan program study saya, beliau juga memang tidak sempat merasakan pendidikan sampai dibangku perkuliahan, namun semangat, motivasi serta sujudnya selalu menjadi doa untuk kesuksesan anak-anaknya.
3. Kepada cinta kasih kedua adik saya, Lutviana Ulfa dan Muhammad Ilham, terimakasih telah memberikan semangat dan dukungannya. Semoga di masa depan, kalian akan lebih berhasil. Teruslah berusaha dan bersemangat.
4. Tidak lupa, saya ucapkan terima kasih kepada Bapak Ainul Yaqin, S.Kom., M.Kom, karena telah sabar dan ikhlas dalam membimbing saya dalam penulisan skripsi ini.
5. Serta ucapan terimakasih banyak kepada teman-teman 20-IF-06, atas semua kebersamaan selama perjalanan perkuliahan ini. Terima kasih atas kebersamaannya, kerjasamanya, dan dukungan yang selalu ada. Semoga tali silaturahmi yang terjalin tetap kokoh, dan dapat terus berbagi pemikiran serta pengalaman dalam segala hal.

KATA PENGANTAR

Bismillahirrahmanirrahim.

Ya Allah, saya bersyukur kepada-MU hari ini. Sebab karena pertolongan-MU lah saya bisa mengerjakan skripsi ini dengan baik. Judul skripsi yang saya usulkan untuk memenuhi syarat kelulusan mata kuliah skripsi di Fakultas Ilmu Komputer, Prodi Informatika, Amikom Yogyakarta adalah "PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI UNTUK ANALISIS SENTIMEN KOMENTAR SHOPEE DI GOOGLE PLAY STORE".

Penulis dengan penuh kesadaran, ingin mengucapkan terima kasih kepada semua orang yang sudah memberikan dukungan dan semangat pada saat pengerjaan skripsi ini. Semoga Allah membalas kebaikan kalian dengan yang terbaik.

Kepada :

Yogyakarta, 22 Januari 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
DAFTAR LAMBANG DAN SINGKATAN	xvii
DAFTAR ISTILAH	xviii
INTISARI	xx
<i>ABSTRACT</i>	xxi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	21
2.2.1 Analisis Sentimen	21
2.2.2 Web Scraping	21
2.2.3 Pre-Processing	21
2.2.4 Google Play Store	23
2.2.5 Multinomial Naïve Bayes	23

2.2.6	Random Forest Classifier.....	24
2.2.7	Logistic Regression.....	24
2.2.8	Support Vector Machine (SVM).....	24
2.2.9	K-Nearest Neighbor	25
2.2.10	Extra Trees Classifier.....	26
2.2.11	Synthetic Minority Oversampling Technique (SMOTE)	26
2.2.12	TF-IDF	26
2.2.13	Confusion Matrix	27
2.2.14	TextBlob	28
BAB III METODE PENELITIAN		29
3.1	Metode Penelitian.....	29
3.2	Alur Penelitian.....	29
3.3	Pengumpulan Data.....	31
3.4	Klasifikasi Algoritma Multinomial Naïve Bayes	32
3.4.1	Klasifikasi Multinomial Naïve Bayes + SMOTE	32
3.4.2	Klasifikasi Multinomial Naïve Bayes Non SMOTE.....	33
3.5	Klasifikasi Algoritma Random Forest Classifier.....	37
3.5.1	Klasifikasi Random Forest Classifier + SMOTE.....	37
3.5.2	Klasifikasi Random Forest Classifier Non SMOTE	38
3.6	Klasifikasi Algoritma Logistic Regression.....	42
3.6.1	Klasifikasi Logistic Regression + SMOTE.....	42
3.6.2	Klasifikasi Logistic Regression Non SMOTE	43
3.7	Klasifikasi Algoritma Support Vector Machine	44
3.7.1	Klasifikasi Support Vector Machine + SMOTE.....	44
3.7.2	Klasifikasi Support Vector Machine Non SMOTE	45
3.8	Klasifikasi Algoritma K-Nearest Neighbors.....	47
3.8.1	Klasifikasi K-Nearest Neighbors + SMOTE	47
3.8.2	Klasifikasi K-Nearest Neighbors Non SMOTE.....	48
3.9	Klasifikasi Algoritma Extra Trees Classifier.....	50
3.9.1	Klasifikasi Extra Trees Classifier + SMOTE.....	50
3.9.2	Klasifikasi Extra Trees Classifier Non SMOTE.....	50

3.10 Alat dan Bahan.....	53
BAB IV HASIL DAN PEMBAHASAN	55
4.1 Hasil Pengambilan Data.....	55
4.2 Hasil Translate Data.....	56
4.3 Pelabelan Data	57
4.3.1 Pelabelan Data Manual	57
4.3.2 Pelabelan Data dengan TextBlob.....	58
4.4 Hasil Preprocessing.....	59
4.4.1 Data Cleaning.....	61
4.4.2 Case Folding	63
4.4.3 Stopword Removal.....	65
4.4.4 Tokenizing	66
4.4.5 Stemming	68
4.5 Exploratory Data Analysis.....	70
4.6 Word Cloud.....	71
4.7 Pembagian Data	73
4.8 Ekstraksi Fitur.....	74
4.9 Synthetic Minority Over-sampling Technique.....	76
4.10 Hasil Klasifikasi.....	78
4.11 Hasil Evaluasi	85
4.11.1 Evaluasi Algoritma Skenario Pelabelan Manual	85
4.11.2 Evaluasi Algoritma Skenario Pelabelan Manual + SMOTE.....	96
4.11.3 Evaluasi Algoritma Skenario Pelabelan TextBlob.....	107
4.11.4 Evaluasi Algoritma Skenario Pelabelan TextBlob + SMOTE.....	119
BAB V PENUTUP	131
5.1 Kesimpulan	131
5.2 Saran	134
REFERENSI	136
LAMPIRAN.....	141

DAFTAR TABEL

Tabel 2.1 Hasil Pengujian Parameter Jumlah Pohon	12
Tabel 2.2 Hasil Pengujian Parameter Kedalaman Pohon	12
Tabel 2.3 Keaslian Penelitian	13
Tabel 2.4 <i>Confusion Matrix</i>	27
Tabel 3.1 Data <i>Training</i> Klasifikasi <i>Multinomial Naïve bayes</i>	33
Tabel 3.2 Jumlah Kemunculan Kata	33
Tabel 3.3 Data <i>Training</i> dan <i>Testing Multinomial Naïve Bayes</i>	36
Tabel 3.4 Data <i>Training</i> dan <i>Testing Random Forest Classifier</i>	39
Tabel 3.5 Bootstrap Sample 1	39
Tabel 3.6 Bootstrap Sample 2	39
Tabel 3.7 Contoh Kalimat <i>Logistic Regression</i>	43
Tabel 3.8 Contoh Kalimat <i>Support Vector Machine</i>	46
Tabel 3.9 Data Training <i>K-Nearest Neighbors</i>	48
Tabel 3.10 Data Testing <i>K-Nearest Neighbors</i>	48
Tabel 3.11 Data <i>Training Extra Trees Classifier</i>	51
Tabel 3.12 Data <i>Testing Extra Trees Classifier</i>	51
Tabel 4.1 Hasil Pengambilan Data	56
Tabel 4.2 Hasil Translate Data	56
Tabel 4.3 Hasil Pelabelan Data Manual	58
Tabel 4.4 Hasil Pelabelan Data dengan <i>TextBlob</i>	59
Tabel 4.5 Hasil Pencarian Kolom Bernilai <i>Null</i> dengan Pelabelan Manual	60
Tabel 4.6 Hasil Pencarian Kolom Bernilai <i>Null</i> dengan Pelabelan <i>TextBlob</i>	60
Tabel 4.7 Hasil Metode <i>Handling Missing Value Ignore Tuple</i>	60
Tabel 4.8 Hasil Data <i>Cleaning</i> dataset bahasa Indonesia	62
Tabel 4.9 Hasil Data <i>Cleaning</i> Dataset Bahasa Inggris	62
Tabel 4.10 Hasil Proses <i>Case Folding</i> Dataset Bahasa Indonesia	64
Tabel 4.11 Hasil <i>Case Folding</i> Dataset Bahasa Inggris	64
Tabel 4.12 Hasil Proses <i>Stopword Removal</i> Dataset Bahasa Indonesia	65
Tabel 4.13 Hasil <i>Stopword Removal</i> Dataset Bahasa Inggris	66

Tabel 4.14 Hasil Proses <i>Tokenizing</i> Bahasa Indonesia.....	67
Tabel 4.15 Hasil <i>Tokenizing</i> Dataset Bahasa Inggris.....	68
Tabel 4.16 Hasil Proses <i>Stemming</i> Bahasa Indonesia.....	69
Tabel 4.17 Hasil <i>Stemming</i> Dataset Bahasa Inggris	70
Tabel 4.18 Skenario Pembagian Data	74
Tabel 4.19 Hasil Pemisahan <i>Term</i>	74
Tabel 4.20 Hasil TF Normalisasi	75
Tabel 4.21 Hasil Penjumlahan DF	75
Tabel 4.22 Hasil Penjumlahan IDF	76
Tabel 4.23 Hasil Perhitungan TF-IDF	76
Tabel 4.24 Hasil Klasifikasi Skenario Pelabelan Manual.....	79
Tabel 4.25 Hasil Klasifikasi Skenario Pelabelan Manual + SMOTE.....	80
Tabel 4.26 Hasil Klasifikasi Skenario Pelabelan <i>TextBlob</i>	82
Tabel 4.27 Hasil Klasifikasi Skenario Pelabelan <i>TextBlob</i> + SMOTE.....	83
Tabel 4.28 Hasil Evaluasi Pelabelan Manual.....	85
Tabel 4.29 Confusion Matrix Multinomial Naïve Bayes.....	86
Tabel 4.30 Confusion Matrix Random Forest Classifier.....	87
Tabel 4.31 <i>Confusion Matrix Logistic Regression</i>	89
Tabel 4.32 <i>Confusion Matrix Support Vector Machine</i>	91
Tabel 4.33 <i>Confusion Matrix K-Nearest Neighbors</i>	93
Tabel 4.34 <i>Confusion Matrix Extra Trees Classifier</i>	94
Tabel 4.35 Hasil Evaluasi Pelabelan Manual + SMOTE.....	96
Tabel 4.36 <i>Confusion Matrix Multinomial Naïve Bayes</i>	97
Tabel 4.37 <i>Confusion Matrix Random Forest Classifier</i>	99
Tabel 4.38 <i>Confusion Matrix Logistic Regression</i>	100
Tabel 4.39 <i>Confusion Matrix Support Vector Machine</i>	102
Tabel 4.40 <i>Confusion Matrix K-Nearest Neighbors</i>	104
Tabel 4.41 <i>Confusion Matrix Extra Trees Classifier</i>	106
Tabel 4.42 Hasil Evaluasi Pelabelan <i>TextBlob</i>	107
Tabel 4.43 <i>Confusion Matrix Multinomial Naïve Bayes</i>	108
Tabel 4.44 <i>Confusion Matrix Random Forest Classifier</i>	110

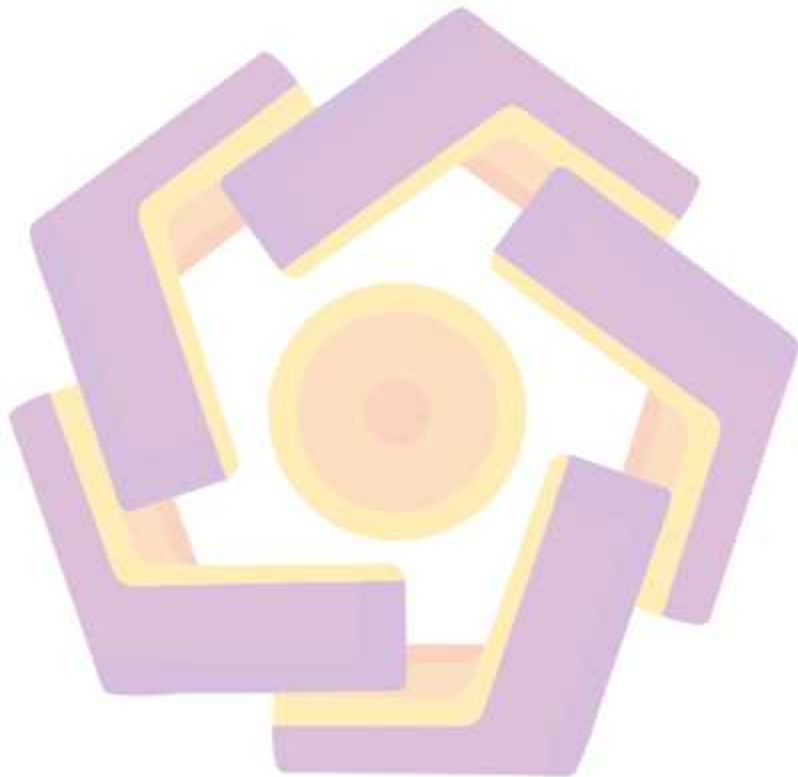
Tabel 4.45 <i>Confusion Matrix Logistic Regression</i>	112
Tabel 4.46 <i>Confusion Matrix Support Vector Machine</i>	113
Tabel 4.47 <i>Confusion Matrix K-Nearest Neighbors</i>	115
Tabel 4.48 <i>Confusion Matrix Extra Trees Classifier</i>	117
Tabel 4.49 Hasil Evaluasi Pelabelan <i>TextBlob</i> + SMOTE	119
Tabel 4.50 <i>Confusion Matrix Multinomial Naïve Bayes</i>	119
Tabel 4.51 <i>Confusion Matrix Random Forest Classifier</i>	121
Tabel 4.52 <i>Confusion Matrix Logistic Regression</i>	123
Tabel 4.53 <i>Confusion Matrix Support Vector Machine</i>	125
Tabel 4.54 <i>Confusion Matrix K-Nearest Neighbors</i>	126
Tabel 4.55 <i>Confusion Matrix Extra Trees Classifier</i>	128



DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	29
Gambar 3.2 Alur Metode <i>Multinomial Naïve Bayes</i> + SMOTE.....	32
Gambar 3.3 Alur Metode <i>Multinomial Naïve Bayes</i> Non SMOTE.....	33
Gambar 3.4 Alur Metode <i>Random Forest Classifier</i> + SMOTE.....	37
Gambar 3.5 Alur Metode <i>Random Forest Classifier</i> Non SMOTE.....	38
Gambar 3.6 Pohon Keputusan <i>Random Forest Classifier</i>	40
Gambar 3.7 Pohon Keputusan <i>Random Forest Classifier</i>	41
Gambar 3.8 Alur Metode <i>Logistic Regression</i> + SMOTE.....	42
Gambar 3.9 Alur Metode <i>Logistic Regression</i> Non SMOTE.....	43
Gambar 3. 10 Alur Metode <i>Support Vector Machine</i> + SMOTE.....	45
Gambar 3. 11 Alur Metode <i>Support Vector Machine</i> Non SMOTE.....	46
Gambar 3.12 Alur Metode <i>K-Nearest Neighbors</i> + SMOTE.....	47
Gambar 3.13 Alur Metode <i>K-Nearest Neighbors</i> Non SMOTE.....	48
Gambar 3. 14 Alur Metode <i>Extra Trees Classifier</i> + SMOTE.....	50
Gambar 3. 15 Alur Metode <i>Extra Trees Classifier</i> Non SMOTE.....	51
Gambar 3.16 Pohon Keputusan <i>Extra Trees Classifier</i>	52
Gambar 4. 1 Gambaran Umum Pengambilan Data.....	55
Gambar 4. 2 Gambaran Umum Pelabelan Data Manual.....	57
Gambar 4. 3 Gambaran Umum Pelabelan Data dengan <i>TextBlob</i>	58
Gambar 4. 4 Gambaran Umum <i>Preprocessing</i>	61
Gambar 4. 5 Gambaran Umum <i>Data Cleaning</i>	61
Gambar 4. 6 Gambaran Umum <i>Case Folding</i>	63
Gambar 4. 7 Gambaran Umum <i>Stopword Removal</i>	65
Gambar 4. 8 Gambaran Umum <i>Tokenizing</i>	67
Gambar 4. 9 Gambaran Umum <i>Stemming</i>	69
Gambar 4. 10 Visualisasi <i>Exploratory Data Analysis</i> Pelabelan Manual.....	70
Gambar 4. 11 Visualisasi <i>Exploratory Data Analysis</i> Pelabelan <i>TextBlob</i>	71
Gambar 4. 12 <i>Word Cloud</i> Sentimen Positif Bahasa Indonesia.....	72
Gambar 4. 13 <i>Word Cloud</i> Sentimen Negatif Bahasa Indonesia.....	72

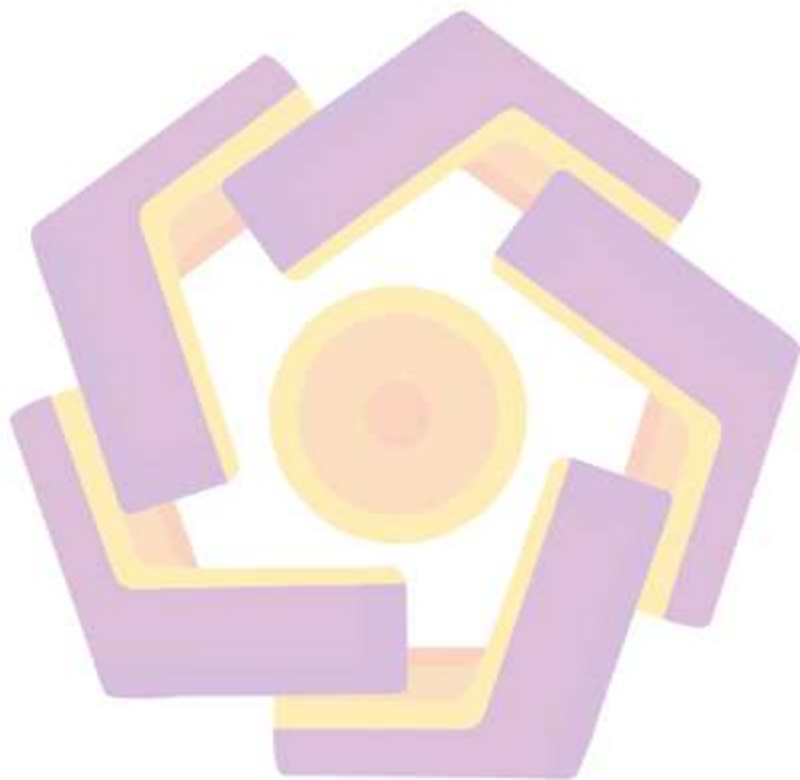
Gambar 4. 14 <i>Word Cloud</i> Sentimen Positif Bahasa Inggris.....	73
Gambar 4. 15 <i>Word Cloud</i> Sentimen Negatif Bahasa Inggris.....	73
Gambar 4. 16 Distribusi <i>Class</i> Metode Pelabelan Manual Sebelum SMOTE	77
Gambar 4. 17 Distribusi <i>Class</i> Metode Pelabelan <i>TextBlob</i> Sebelum SMOTE	77
Gambar 4. 18 Distribusi <i>Class</i> Setelah Menerapkan SMOTE.....	78



DAFTAR LAMPIRAN

Lampiran 1 <i>Script Scraping Data</i>	141
Lampiran 2 <i>Script Translate Data</i>	141
Lampiran 3 <i>Script Pelabelan Manual</i>	142
Lampiran 4 <i>Script Pelabelan TextBlob</i>	142
Lampiran 5 <i>Script Untuk Mencari Kolom Bernilai Null</i>	143
Lampiran 6 <i>Script Handling Missing Value Ignore Tuple</i>	143
Lampiran 7 <i>Script Data Cleaning dan Case Folding</i>	143
Lampiran 8 <i>Script Stopword Removal Bahasa Indonesia</i>	143
Lampiran 9 <i>Script Stopword Removal Bahasa Inggris</i>	143
Lampiran 10 <i>Script Tokenizing</i>	144
Lampiran 11 <i>Script Stemming</i>	144
Lampiran 12 <i>Script Menyimpan Data Hasil Preprocessing</i>	145
Lampiran 13 <i>Script EDA Menghitung Jumlah Baris Nilai Label</i>	145
Lampiran 14 <i>Script EDA untuk Membuat Diagram Bar</i>	145
Lampiran 15 <i>Script WordCloud Sentimen Positif</i>	145
Lampiran 16 <i>Script WordCloud Sentimen Negatif</i>	146
Lampiran 17 <i>Script Pembagian Data</i>	146
Lampiran 18 <i>Script TF-IDF</i>	146
Lampiran 19 <i>Script Mengaplikasikan TF-IDF pada Data Training Testing</i>	147
Lampiran 20 <i>Script untuk Mengaplikasikan SMOTE pada Data Training</i>	147
Lampiran 21 <i>Script Klasifikasi Multinomial Naïve Bayes dengan SMOTE</i>	147
Lampiran 22 <i>Script Klasifikasi Multinomial Naïve Bayes tanpa SMOTE</i>	148
Lampiran 23 <i>Script Klasifikasi Random Forest dengan SMOTE</i>	149
Lampiran 24 <i>Script Klasifikasi Random Forest tanpa SMOTE</i>	150
Lampiran 25 <i>Script Klasifikasi Logistic Regression dengan SMOTE</i>	150
Lampiran 26 <i>Script Klasifikasi Logistic Regression tanpa SMOTE</i>	151
Lampiran 27 <i>Script Klasifikasi SVM Kernel Linear dengan SMOTE</i>	151
Lampiran 28 <i>Script Klasifikasi SVM Kernel Linear tanpa SMOTE</i>	152

Lampiran 29 <i>Script</i> Klasifikasi KNN dengan SMOTE	153
Lampiran 30 <i>Script</i> Klasifikasi KNN tanpa SMOTE	153
Lampiran 31 <i>Script</i> Klasifikasi <i>Extra Trees</i> dengan SMOTE	154
Lampiran 32 <i>Script</i> Klasifikasi <i>Extra Trees</i> tanpa SMOTE	155



DAFTAR LAMBANG DAN SINGKATAN

SMOTE	Synthetic Minority Over-sampling Technique
TF-IDF	Term Frequency-Inverse Document Frequency
CPD	Categorical Proportional Difference
SVM	Support Vector Machines
AUC	Area Under the Curve
RBF	Radial Basis Function Kernel
KDD	Knowledge Discovery In Database
KNN	K-Nearest Neighbors
BMKG	Badan Meteorologi, Klimatologi, dan Geofisika
NLP	Natural Language Processing
AI	Artificial Intelligence
TF	Term Frequency
IDF	Inverse Document Frequency
TP	True Positif
TN	True Negatif
FP	False Positif
FN	False Negatif
CSV	Comma-Separated Values
EDA	Exploratory Data Analysis
Scikit-Learn	Scientific Kit for Machine Learning
Google Colab	Google Colaboratory
NLTK	Natural Language Toolkit

DAFTAR ISTILAH



Sentimen	Perasaan positif, negatif, atau netral yang terkandung dalam sebuah teks.
Polaritas	melihat sejauh mana sentimen dalam suatu teks cenderung positif atau negatif.
Opini	Pendapat seseorang yang mencerminkan sentimen terhadap suatu isu dan bersifat subjektif.
Klasifikasi Sentimen	Pengelompokkan teks kedalam kelas sentimen, seperti sentimen positif, negatif, atau netral.
Natural Language Processing	membuat komputer dapat berpikir dan bekerja mirip dengan manusia agar dapat memproses dan memahami suatu teks.
Tokenizing	memisahkan kalimat yang panjang menjadi perkata atau yang lebih dikenal dengan sebutan token.
Preprocessing	mengolah data mentah, memperbaiki data mentah, menyeleksi data mentah dan memastikan bahwa data mentah tersebut telah siap digunakan untuk proses analisis.
Confusion Matrix	tabel yang sering dipakai untuk mengevaluasi kinerja dari klasifikasi <i>machine learning</i> yang dapat menampilkan dan membandingkan data yang sebenarnya dengan nilai hasil prediksi model.
Machine Learning	Pendekatan analisis sentimen yang memanfaatkan algoritma pembelajaran mesin untuk mengetahui pola-pola dan memprediksi sentimen.



Accuracy	Matrix evaluasi yang digunakan untuk melihat seberapa baik model analisis sentimen bisa memprediksi sentimen dengan benar.
Stopword Removal	Menghapus kata yang tidak berguna, dan dianggap tidak bermakna, kata tidak baku dan tidak berpengaruh pada analisis.
Imbalanced Data	Kondisi dimana jumlah sentimen positif dan sentimen negatif tidak seimbang datanya.
F1-score	Matrix evaluasi yang bekerja dengan menggabungkan hasil precision dan hasil recall, yang tujuannya untuk memberikan gambaran tentang kinerja model ketika ada imbalanced data.
Precision	Melihat sejauh mana prediksi positif dari suatu model ialah benar.
Recall	Melihat sejauh mana model berhasil bisa mendeteksi semua instance positif yang seharusnya.
Data Cleaning	Membersihkan dan menghapus semua karakter selain huruf (a-z dan A-Z) seperti emoji, angka, tanda baca dan karakter khusus lainnya.
Case Folding	Menyeragamkan huruf kapital menjadi <i>lowercase</i> semua.
Stemming	Tahapan untuk melakukan transformasi kata dengan memakai kata dasarnya dan menghilangkan kata imbuhan, awalan dan akhiran.
Overfitting	Kondisi dimana model dapat memprediksi data training secara akurat tapi tidak untuk data baru.

INTISARI

Shopee merupakan aplikasi belanja online terbesar di Indonesia dan berhasil meraih peringkat pertama dalam kategori *marketplace* pada kuartal II tahun 2023. Akan tetapi, semua itu tidak dapat digunakan sebagai tolak ukur kepuasan pengguna. Dalam konteks ini, kepuasan pengguna hanya dapat diukur dari komentar dan rating yang diberikan oleh pengguna aplikasi Shopee di Google Play Store. Oleh sebab itu, diperlukan analisis sentimen terhadap komentar aplikasi Shopee, yang bertujuan untuk melihat komentar masyarakat tentang aplikasi Shopee di Google Play Store cenderung positif atau negatif. Untuk melaksanakan analisis sentimen terhadap komentar tersebut, data yang digunakan adalah data komentar dan rating pengguna Shopee di Google Play Store, yang berjumlah 1500 data. Terdapat empat skenario dalam tahap klasifikasi. Pertama, metode pelabelan manual. Kedua, metode pelabelan manual + SMOTE pada implementasi algoritma klasifikasi. Ketiga, pelabelan menggunakan *TextBlob*, dan keempat, pelabelan *TextBlob* + SMOTE pada implementasi algoritma klasifikasi, yaitu *Multinomial Naïve Bayes*, *Random Forest Classifier*, *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Extra Trees Classifier*, untuk membandingkan kinerja semua algoritma tersebut dan menentukan algoritma mana yang memberikan akurasi tertinggi. Pengujian dilakukan menggunakan *confusion matrix* untuk mengevaluasi kinerja dari masing-masing algoritma klasifikasi. Hasil pengujian menunjukkan bahwa algoritma *K-Nearest Neighbors* pada skenario ketiga mencapai akurasi tertinggi, yaitu sebesar 90,35%. Temuan dari penelitian ini dapat menjadi panduan bagi pihak manajemen Shopee untuk mengembangkan aplikasinya dan meningkatkan kualitas layanan.

Kata kunci: *Natural Language Processing*, Analisis Sentimen, Shopee, Google Play Store, SMOTE.

ABSTRACT

Shopee is the largest online shopping application in Indonesia and was ranked first in the marketplace category in the second quarter of 2023. However, all of that cannot be used as a measure of user satisfaction. In this context, user satisfaction can only be measured from the comments and ratings given by Shopee app users on the Google Play Store. Therefore, a sentiment analysis of Shopee app comments is needed, which aims to see whether people's comments about the Shopee app on the Google Play Store tend to be positive or negative. To carry out sentiment analysis of these comments, the data used is data on comments and ratings of Shopee users on the Google Play Store, which amounts to 1500 data. There are four scenarios in the classification stage. First, the manual labeling method. Second, the manual + SMOTE labeling method in the implementation of the classification algorithm. Third, labeling using TextBlob, and fourth, TextBlob + SMOTE labeling on the implementation of classification algorithms, namely Multinomial Naive Bayes, Random Forest Classifier, Logistic Regression, Support Vector Machine, K-Nearest Neighbors, and Extra Trees Classifier, to compare the performance of all these algorithms and determine which algorithm provides the highest accuracy. Tests were conducted using confusion matrix to evaluate the performance of each classification algorithm. The test results showed that the K-Nearest Neighbors algorithm in the third scenario achieved the highest accuracy of 90.35%. The findings from this research can be a guide for Shopee management to develop its application and improve service quality.

Keyword: Natural Language Processing, Sentiment Analysis, Shopee, Google Play Store, SMOTE.