

**PERBANDINGAN LABELING DALAM ANALISIS SENTIMEN  
KOMENTAR PENGGUNA APLIKASI GOJEK DI PLATFORM  
GOOGLE PLAY STORE**

**SKRIPSI**

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



disusun oleh

**ADI SAPRIANTO**

**20.11.3614**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2024**

**PERBANDINGAN LABELING DALAM ANALISIS SENTIMEN  
KOMENTAR PENGGUNA APLIKASI GOJEK DI PLATFORM  
GOOGLE PLAY STORE**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**ADI SAPRIANTO**

**20.11.3614**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN LABELING DALAM ANALISIS SENTIMEN  
KOMENTAR PENGGUNA APLIKASI GOJEK DI PLATFORM  
GOOGLE PLAY STORE**

yang disusun dan diajukan oleh

**Adi Saprianto**

**20.11.3614**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 22 Agustus 2024

Dosen Pembimbing,



**Whindha Mega Pradnya Duhita, M.Kom**

**NIK. 190302185**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**PERBANDINGAN LABELING DALAM ANALISIS SENTIMEN**  
**KOMENTAR PENGGUNA APLIKASI GOJEK DI PLATFORM**  
**GOOGLE PLAY STORE**

yang disusun dan diajukan oleh

**Adi Saprianto**

20.11.3614

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 22 Agustus 2024

**Susunan Dewan Penguji**

**Nama Penguji**

**Windha Mega Pradnya, M.Kom.**

**NIK. 190302185**

**Anna Baita, M.Kom.**

**NIK. 190302290**

**Rizqi Sukma Kharisma, M.kom.**

**NIK. 190302215**

**Tanda Tangan**



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

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, M.Kom., Ph.D.**

**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Adi Saprianto  
NIM : 20.11.3614

Menyatakan bahwa Skripsi dengan judul berikut:

### PERBANDINGAN LABELING DALAM ANALISIS SENTIMEN KOMENTAR PENGGUNA APLIKASI GOJEK DI PLATFORM GOOGLE PLAY STORE

Dosen Pembimbing : Windha Mega Pradnya Dhuhita, 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 Agustus 2024

Yang Menyatakan,



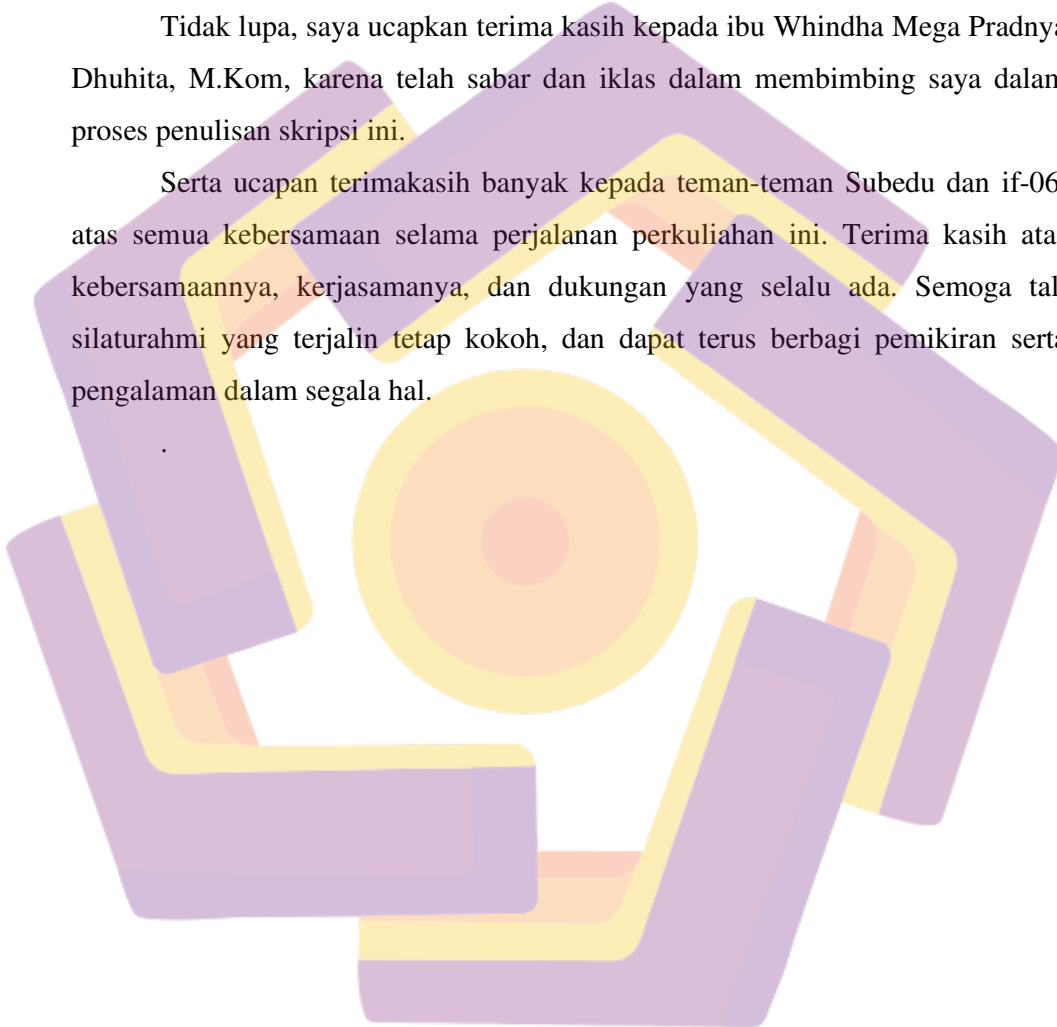
Adi Saprianto

## HALAMAN PERSEMBAHAN

Skripsi ini disusun dan dipersembahkan dengan penuh rasa syukur dan cinta kepada keluarga tercinta, yang senantiasa memberikan dukungan tidak terhingga selama perjalanan pendidikan ini.

Tidak lupa, saya ucapkan terima kasih kepada ibu Whindha Mega Pradnya Duhita, M.Kom, karena telah sabar dan ikhlas dalam membimbing saya dalam proses penulisan skripsi ini.

Serta ucapan terimakasih banyak kepada teman-teman Subedu dan 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

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "PERBANDINGAN LABELING DALAM ANALISIS SENTIMEN KOMENTAR PENGGUNA APLIKASI GOJEK DI PLATFORM GOOGLE PLAY STORE " ini dengan baik. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak prof. Dr. M. Suyanto, M.M. ,Rektor Universitas Amikom Yogyakarta., yang telah memberikan kesempatan dan fasilitas yang diperlukan dalam penyusunan skripsi ini.
2. Bapak Hanif Al Fatta, M.Kom., Ph.D., Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, yang telah memberikan dukungan moral dan material dalam proses penyusunan skripsi ini.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku ketua program Studi S1 Informatika Universitas Amikom Yogyakarta. Dan juga selaku dosen pembimbing saya, yang telah dengan sabar membimbing dan memberikan arahan selama penyusunan skripsi ini hingga selesai.
4. Orang Tua dan Keluarga Tercinta, atas segala doa, dukungan, dan kasih sayang yang senantiasa mengiringi langkah penulis.
5. Sahabat-Sahabat dan seluruh pihak yang telah membantu dan memberikan dukungan dalam penyelesaian skripsi ini, baik secara langsung maupun tidak langsung.

penulis sangat mengharapkan kritik dan saran yang membangun demi kesempurnaan skripsi ini di masa mendatang. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pembaca serta dapat menjadi kontribusi yang bermanfaat dalam bidang ilmu yang ditekuni.

Yogyakarta, 22 Agustus 2024

Penulis

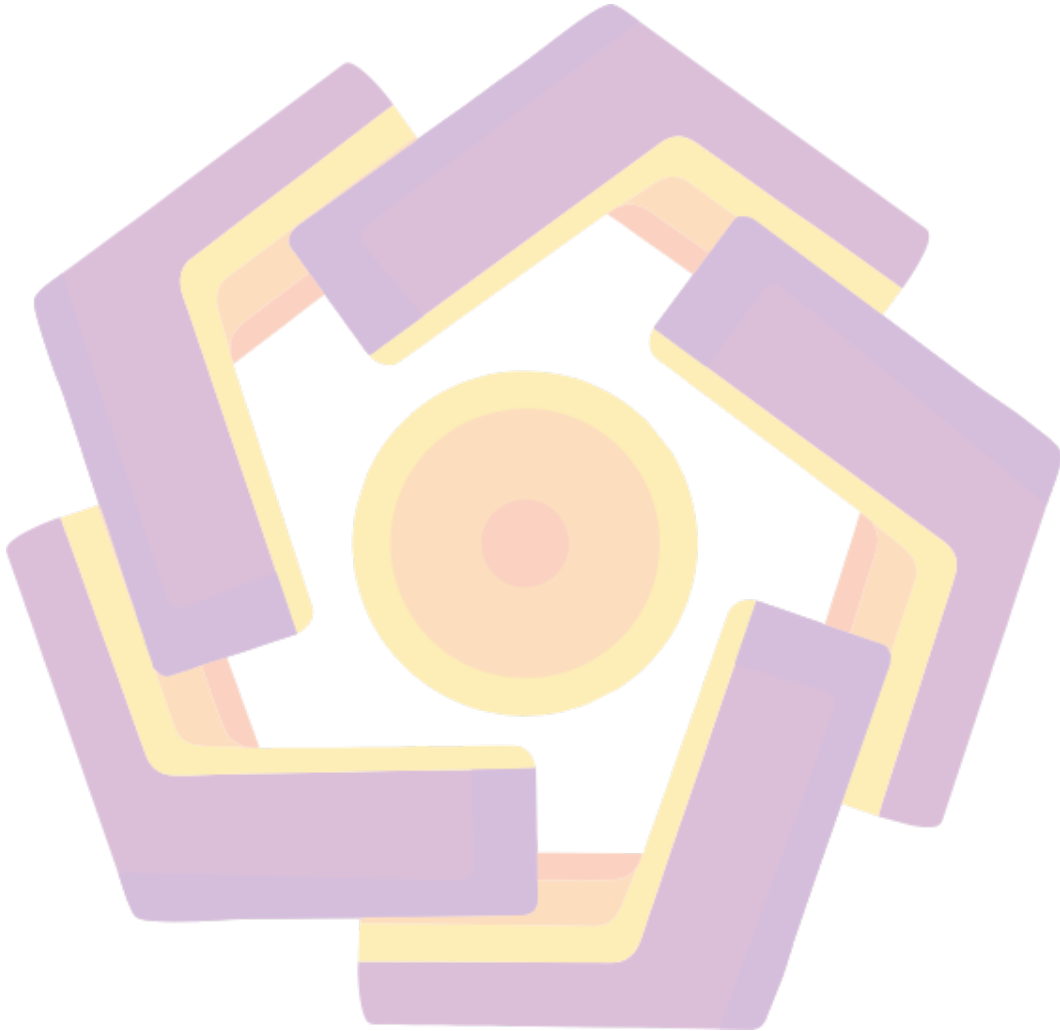
## DAFTAR ISI

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 .....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN .....	xiv
DAFTAR ISTILAH.....	xv
INTISARI .....	xviii
<i>ABSTRACT</i> .....	xix
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 .....	15
BAB III METODE PENELITIAN .....	22
3.1 Objek Penelitian.....	22
3.2 Alur Penelitian .....	22



3.3	Alat dan Bahan.....	24
BAB IV HASIL DAN PEMBAHASAN .....		26
4.1	Hasil Pengambilan Data.....	26
4.2	Hasil Translate Data.....	27
4.3	Pelabelan Data .....	28
4.3.1	Pelabelan Data Manual .....	28
4.3.2	Pelabelan Data TextBlob .....	29
4.4	Hasil Preprocessing.....	31
4.4.1	Data Cleaning.....	32
4.4.2	Case Folding .....	34
4.4.3	Stopword Removal.....	36
4.4.4	Tokenizing .....	37
4.4.5	Stemming .....	39
4.5	Exploratory Data Analysis.....	40
4.6	Word Cloud.....	42
4.7	Pembagian Data .....	45
4.8	Ekstraksi Fitur.....	46
4.9	Synthetic Minority Over-sampling Technique .....	48
4.10	Hasil Klasifikasi.....	50
4.10.1	Evaluasi Kernel RBF dengan Pelabelan Manual .....	52
4.10.2	Evaluasi Kernel Polynomial dengan Pelabelan Manual .....	54
4.10.3	Evaluasi Kernel Linear dengan Pelabelan Manual .....	56
4.10.4	Evaluasi Kernel Sigmoid dengan Pelabelan Manual .....	58
4.10.5	Evaluasi Kernel RBF dengan Pelabelan TextBlob .....	60
4.10.6	Evaluasi Kernel Polynomial dengan Pelabelan TextBlob .....	62
4.10.7	Evaluasi Kernel Linear dengan Pelabelan TextBlob .....	64
4.10.8	Evaluasi Kernel Sigmoid dengan Pelabelan TextBlob .....	66

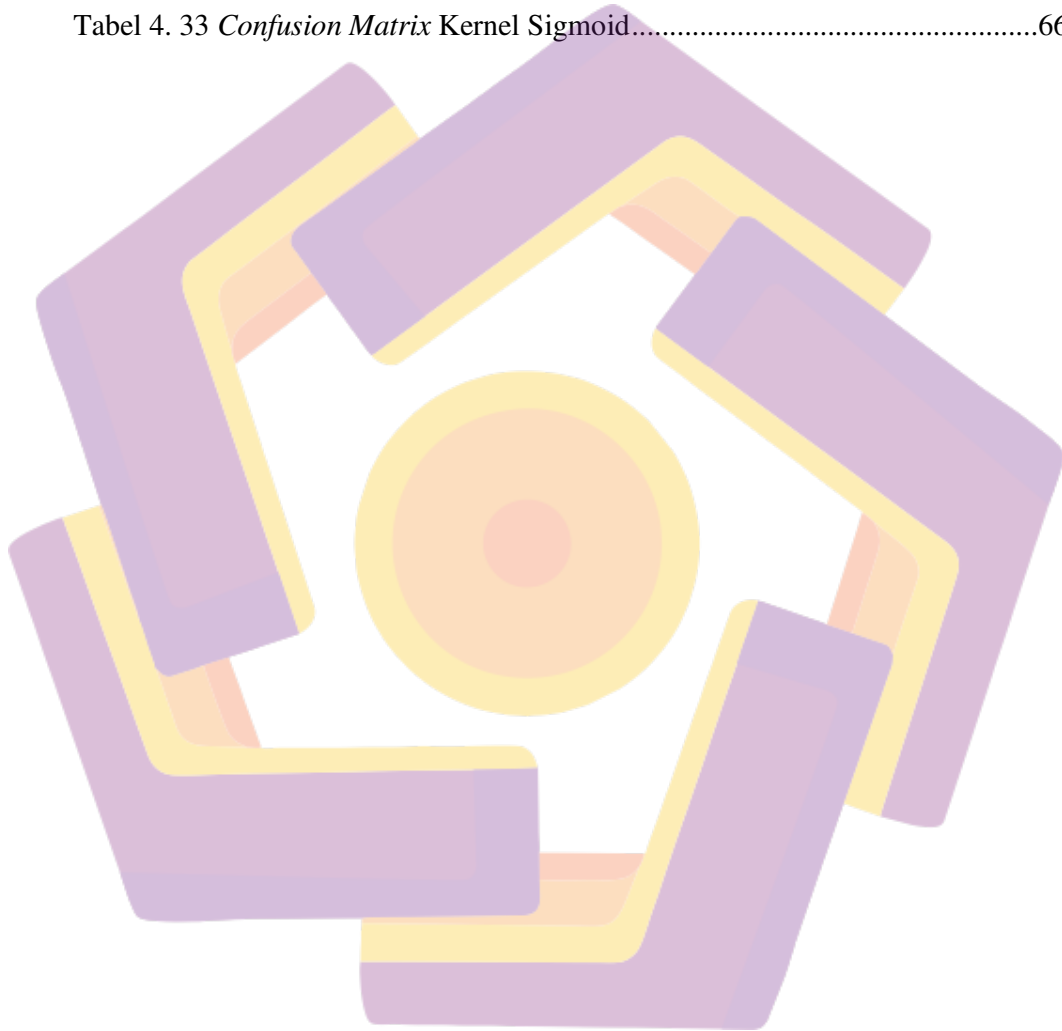
BAB V PENUTUP .....	69
5.1 Kesimpulan .....	69
5.2 Saran .....	69
REFRENSI.....	70
LAMPIRAN.....	73



## DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian .....	10
Tabel 2. 2 Confusion matrix .....	20
Tabel 4. 1 Hasil Pengambilan Data.....	27
Tabel 4. 2 Hasil Translate Data.....	27
Tabel 4. 3 Hasil Pelabelan Manual .....	28
Tabel 4. 4 Hasil Pelabelan <i>TextBlob</i> .....	30
Tabel 4. 5 hasil pencarian baris data <i>Null</i> pada pelabelan manual .....	31
Tabel 4. 6 Hasil Pencarian Baris data <i>Null</i> Pada Pelabelan <i>TextBlob</i> .....	31
Tabel 4. 7 Hasil <i>handling missing value ignore tuple</i> .....	31
Tabel 4. 8 Hasil Cleaning Data Bahasa Indonesia.....	32
Tabel 4. 9 Hasil Cleaning Data Bahasa Inggris .....	33
Tabel 4. 10 Hasil <i>Case Folding</i> Dataset Bahasa Indonesia .....	34
Tabel 4. 11 Hasil <i>Case Folding</i> Dataset Bahasa Inggris .....	35
Tabel 4. 12 Hasil Proses Stopword Removal Dataset Bahasa Indonesia .....	36
Tabel 4. 13 Hasil Proses <i>Stopword Removal</i> Dataset Bahasa Inggris .....	37
Tabel 4. 14 Hasil <i>Tokenizing</i> dataset Bahasa Indonesia .....	38
Tabel 4. 15 Hasil <i>Tokenizing</i> dataset Bahasa Inggris.....	38
Tabel 4. 16 Hasil Proses <i>Stemming</i> Bahasa Indonesia.....	39
Tabel 4. 17 Hasil Proses <i>Stemming</i> Bahasa Inggris.....	40
Tabel 4. 18 Skenario Pembagian Data .....	45
Tabel 4. 19 Hasil Pemisahan <i>Term</i> .....	46
Tabel 4. 20 Hasil TF Normalisasi .....	46
Tabel 4. 21 Hasil penjumlahan <i>term</i> .....	47
Tabel 4. 22 Hasil dari perhitungan nilai IDF .....	47
Tabel 4. 23 Hasil Perhitungan TF-IDF .....	47
Tabel 4. 24 Hasil Klasifikasi Skenario Pelabelan Manual.....	50
Tabel 4. 25 Hasil Klasifikasi Skenario Pelabelan <i>TextBlob</i> .....	51
Tabel 4. 26 <i>Confusion Matrix</i> kernel RBF.....	52
Tabel 4. 27 <i>Confusion Matrix</i> kernel Polynomial.....	54

Tabel 4. 28 <i>Confusion Matrix</i> kernel Linear.....	56
Tabel 4. 29 <i>Confusion Matrix</i> kernel Sigmoid.....	58
Tabel 4. 30 <i>Confusion Matrix</i> kernel RBF.....	60
Tabel 4. 31 <i>Confusion Matrix</i> Kernel Polynomial.....	62
Tabel 4. 32 <i>Confusion Matrix</i> Kernel Linear.....	64
Tabel 4. 33 <i>Confusion Matrix</i> Kernel Sigmoid.....	66




## DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian .....	22
Gambar 4. 1 Proses Pengambilan Data.....	26
Gambar 4. 2 Cara Kerja pelabelan TextBlob.....	29
Gambar 4. 3 Alur Preprocessing .....	32
Gambar 4. 4 Visualisasi <i>Exploratory Data Analysis</i> Pelabelan Manual.....	41
Gambar 4. 5 Visualisasi <i>Exploratory Data Analysis</i> Pelabelan <i>TextBlob</i> .....	42
Gambar 4. 6 <i>word cloud</i> Bahasa Indonesia sentimen Positif.....	43
Gambar 4. 7 <i>word cloud</i> Bahasa Indonesia sentimen Negatif .....	43
Gambar 4. 8 <i>word cloud</i> Bahasa Inggris sentimen Positif.....	44
Gambar 4. 9 <i>word cloud</i> Bahasa Inggris sentimen Negatif .....	45
Gambar 4. 10 Hasil distribusi Pelabelan Manual sebelum SMOTE.....	48
Gambar 4. 11 Hasil distribusi Pelabelan <i>TextBlob</i> Sebelum SMOTE.....	49
Gambar 4. 12 Distribusi <i>Class</i> Setelah Menerapkan SMOTE.....	49

## DAFTAR LAMPIRAN


Lampiran 1 <i>Script Scraping Data</i> .....	73
Lampiran 2 <i>Script Translate Data</i> .....	73
Lampiran 3 <i>Script Pelabelan TextBlob</i> .....	74
Lampiran 4 <i>Script Untuk Mencari Kolom Bernilai Null</i> .....	74
Lampiran 5 <i>Script Handling Missing Value Ignore Tuple</i> .....	74
Lampiran 6 <i>Script Data Cleaning dan Case Folding</i> .....	74
Lampiran 7 <i>Script Stopword Removal Bahasa Indonesia</i> .....	75
Lampiran 8 <i>Script Stopword Removal Bahasa Inggris</i> .....	75
Lampiran 9 <i>Script Tokenizing</i> .....	75
Lampiran 10 <i>Script Stemming</i> .....	75
Lampiran 11 <i>Script Menyimpan Data Hasil Preprocessing</i> .....	77
Lampiran 12 <i>Script EDA Menghitung Jumlah Baris Nilai Label</i> .....	77
Lampiran 13 <i>Script EDA untuk Membuat Diagram Bar</i> .....	77
Lampiran 14 <i>Script WordCloud Sentimen Positif</i> .....	77
Lampiran 15 <i>Script WordCloud Sentimen Negatif</i> .....	78
Lampiran 16 <i>Script Pembagian Data</i> .....	78
Lampiran 17 <i>Script TF-IDF</i> .....	78
Lampiran 18 <i>Script Mengaplikasikan TF-IDF pada Data Training Testing</i> .....	79
Lampiran 19 <i>Script untuk Mengaplikasikan SMOTE pada Data Training</i> .....	79
Lampiran 20 <i>script klasifikasi kernel RBF</i> .....	79
Lampiran 21 <i>script klasifikasi kernel Polynomial</i> .....	80
Lampiran 22 <i>script klasifikasi kernel Linear</i> .....	81
Lampiran 23 <i>script klasifikasi kernel Sigmoid</i> .....	82

## DAFTAR LAMBANG DAN SINGKATAN



SMOTE	Synthetic Minority Over-sampling Technique
TF-IDF	Term Frequency-Inverse Document Frequency
SVM	Support Vector Machines
RBF	Radial Basis Function Kernel
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

## DAFTAR ISTILAH



Sentimen	Perasaan positif, negatif, atau netral yang terkandung dalam sebuah teks.
term	Jumlah kemunculan istilah
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

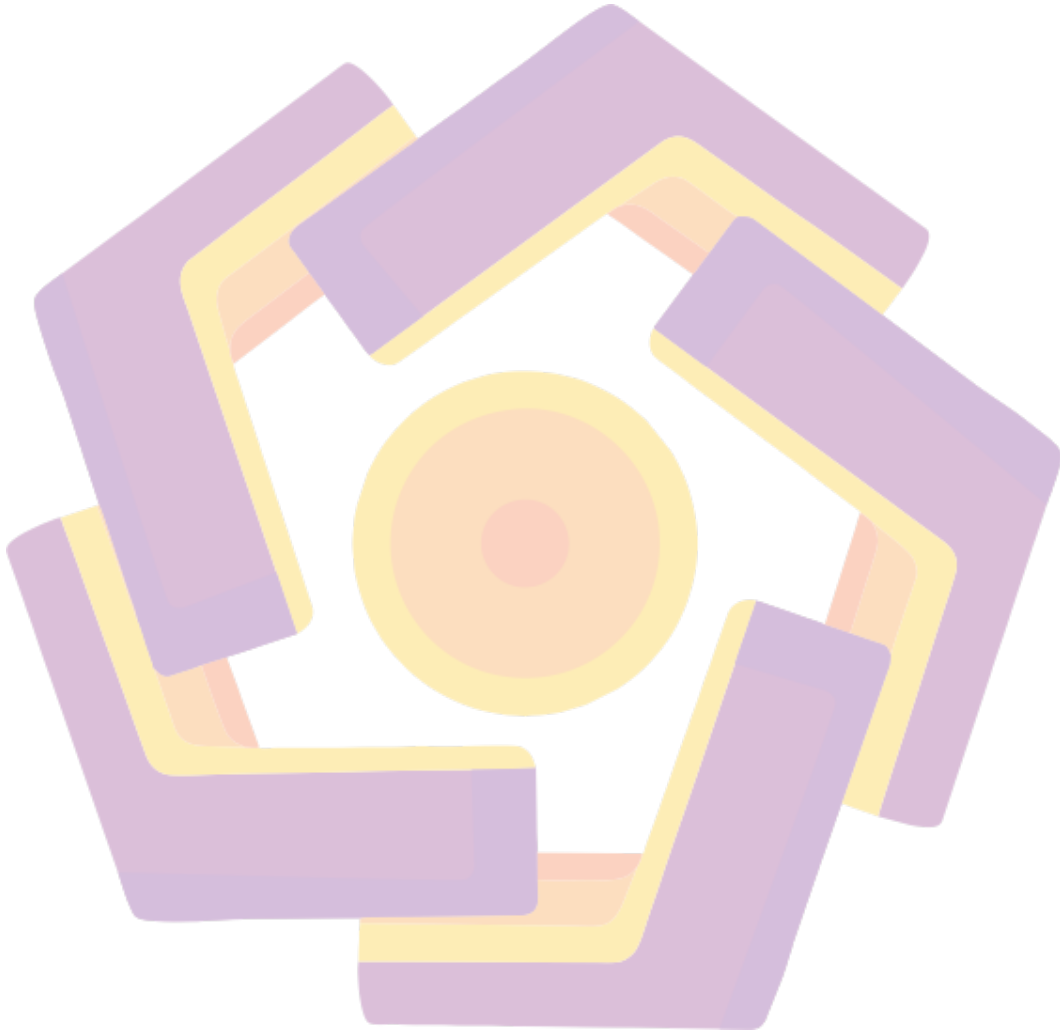
Membersihkan dan menghapus semua karakter selain huruf (a-z dan A-Z) seperti emoji, angka, tanda baca dan karakter khusus lainnya.

Case Folding

Menyeragamkan huruf kapital menjadi *lowercase* semua.

## Stemming

Tahapan untuk melakukan transformasi kata dengan memakai kata dasarnya dan menghilangkan kata imbuhan, awalan dan akhiran.



## INTISARI

Gojek merupakan aplikasi ojek online yang didirikan oleh Nadiem Makarim pada tahun 2010. Gojek berada pada urutan pertama aplikasi transportasi online terbanyak di unduh di RI 2023, dengan rata-rata 957 ribu unduhan perbulan dari pengguna smartphone. hal tersebut tidak bisa dijadikan tolak ukur untuk menilai kepuasan pengguna. Pengguna menyampaikan semua keluhan di kolom komentar yang disediakan oleh Google Play Store. diperlukan penelitian yang bertujuan untuk mengkategorikan komentar tersebut menjadi komentar positif dan negatif. Salah satu cara yang dapat digunakan untuk mengkategorikan komentar tersebut adalah dengan menerapkan analisis sentimen pada komentar aplikasi Gojek di Google Play Store. Dalam penelitian ini menggunakan kernel trick dari algoritma Support Vector Machine. Kernel digunakan untuk menganalisis sentimen dengan membandingkan pelabelan manual dengan pelabelan TextBlob. menggunakan metode scraping untuk mengumpulkan dataset sebanyak 1500 data komentar pengguna aplikasi Gojek. Dataset tersebut kemudian di labeli menjadi komentar positif dan komentar negatif supaya lebih mudah dalam melakukan preproceasing. Tahap selanjutnya melakukan EDA (Exploratory Data Analysis), word cloud, pembagian data, Extrasi fitur TF-IDF, dan SMOTE (Synthetic minority oversampling technique). Hasil dari pengujian yaitu pada metode Pelabelan TextBlob, kernel dengan akurasi terbesar adalah kernel Linear sebesar 85,47%, Sedangkan pada pelabelan manual nilai akurasi terbesar pada kernel Linear dengan akurasi sebesar 83,38%

**Kata kunci:** Gojek, sentimen, SVM, kernel, SMOTE.

## **ABSTRACT**

*Nadiem Makarim founded the online Gojek application in 2010. Gojek ranks first among the most downloaded online transportation apps in RI 2023, with an average of 957 thousand downloads per month from smartphone users. You cannot use it as a metric to gauge user satisfaction. Users submit all complaints in the comment columns provided by the Google Play Store. Research is required to categorize such comments as positive or negative. To categorize such comments, apply sentimental analysis to the Gojek app's comments in the Google Play Store. This study uses a trick kernel from the Support Vector Machine algorithm. We use the kernel to analyze sentiment by comparing manual labeling with TextBlob labeling. It uses scraping methods to collect 1500 user comment data from Gojek applications. We then label the data sets into positive and negative comments to facilitate preprocessing. Next steps are EDA (Exploratory Data Analysis), word cloud, data splitting, TF-IDF feature extraction, and SMOTE. (Synthetic Minority Oversampling Technique). The result of the test is that in the TextBlob Labeling method, the linear kernel has the highest accuracy at 85.47%, whereas in manual labeling, the most accurate value is a linear nucleus at 83.38%.*

**Keyword:** *Gojek, sentiment, SVM, kernel, SMOTE*