

**PREDIKSI KEBERLANJUTAN GOJEK DI PASAR
TRANSPORTASI ONLINE INDONESIA MENGGUNAKAN
ANALISIS SENTIMEN DENGAN ALGORITMA NAÏVE
BAYES, KNN, DAN SVM**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Skripsi

Leonard Agust Laga Lajar

00000043809

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

**PREDIKSI KEBERLANJUTAN GOJEK DI PASAR
TRANSPORTASI ONLINE INDONESIA MENGGUNAKAN
ANALISIS SENTIMEN DENGAN ALGORITMA NAÏVE
BAYES, KNN, DAN SVM**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom.)

Leonard Agust Laga Lajar

00000043809

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2024

i

Prediksi Keberlanjutan Gojek ..., Leonard Agust Laga Lajar, Universitas
Multimedia Nusantara

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Leonard Agust Laga Lajar

Nomor Induk Mahasiswa : 00000043809

Program Studi : Sistem Informasi

Skripsi dengan judul:

Prediksi Keberlanjutan Gojek di Pasar Transportasi Online Indonesia Menggunakan Analisis Sentimen Dengan Algoritma Naive Bayes, KNN, dan SVM

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 29 November 2024



(Leonard Agust Laga Lajar)

HALAMAN PENGESAHAN

Skripsi dengan judul

Prediksi Keberlanjutan Gojek di Pasar Transportasi Online Indonesia
Menggunakan Analisis Sentimen Dengan Algoritma Naive Bayes, KNN, dan
SVM

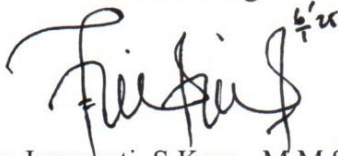
Oleh

Nama : Leonard Agust Laga Lajar
NIM : 00000043809
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Jumat, 6 Desember 2024
Pukul 13.00 s.d 15.00 dan dinyatakan
LULUS

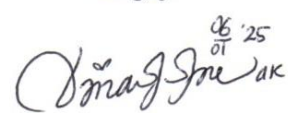
Dengan susunan penguji sebagai berikut.

Ketua Sidang



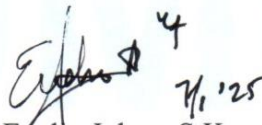
Dr. Irmawati, S.Kom., M.M.S.I.
NIDN: 0805097703/NIK: 081431

Penguji



Dinar Ajeng Kristiyanti, S.Kom., M.Kom.
NIDN: 0330128801/NIK: 079159

Pembimbing



Monika Evelyn Johan, S.Kom.,
M.M.S.I.

NIDN: 0327059501/NIK: 071281

Ketua Program Studi Sistem Informasi



Ririn Ikana Desanti S.Kom., M.Kom.

iii

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Leonard Agust Laga Lajar
NIM : 00000043809
Program Studi : Sistem Informasi
Jenjang : D3/S1/S2* (pilih salah satu)
Judul Karya Ilmiah : Prediksi Keberlanjutan Gojek di Pasar Transportasi Online Indonesia Menggunakan Analisis Sentimen Dengan Algoritma Naive Bayes, KNN, dan SVM Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- Lainnya, pilih salah satu:
- Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 29 November 2024



(Leonard Agust Laga Lajar)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.\

KATA PENGANTAR

Mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, M.A, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi, Universitas Multimedia Nusantara.
4. Ibu Monika Evelin Johan, S.Kom., M.M.S.I., selaku Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi sehingga terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat bermanfaat bagi pembaca dan dikembangkan terus menerus sehingga ilmu yang ada dapat terus berkembang.

Tangerang, 29 November 2024



(Leonard Agust Laga Lajar)
UNIVERSITAS
MULTIMEDIA
NUSANTARA

PREDIKSI KEBERLANJUTAN GOJEK DI PASAR TRANSPORTASI ONLINE INDONESIA MENGGUNAKAN ANALISIS SENTIMEN DENGAN ALGORITMA NAÏVE BAYES, KNN, DAN SVM

ABSTRAK

Transportasi online menjadi kebutuhan penting dalam kehidupan sehari-hari masyarakat Indonesia, dengan Gojek sebagai pemimpin pasar. Namun, sejak 2022 hingga 2023, jumlah unduhan aplikasi Gojek terus menurun. Hal ini menimbulkan kekhawatiran terkait keberlanjutan Gojek di pasar transportasi online. Penelitian ini bertujuan untuk memprediksi keberlanjutan Gojek dengan menganalisis sentimen pengguna berdasarkan ulasan di *Google Play Store*.

Metode yang digunakan adalah analisis sentimen menggunakan algoritma *Naive Bayes*, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM). Penelitian ini mengadopsi framework *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang mencakup tahap pengumpulan data, pemrosesan data, pemodelan, evaluasi, dan implementasi. Data ulasan pengguna di *Google Play Store* diproses untuk menghasilkan klasifikasi sentimen positif, negatif, atau netral.

Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine*(SVM) memberikan performa terbaik dalam klasifikasi sentimen dengan tingkat akurasi sebesar 89,65%, *precision* 92%, *recall* 93%, dan *f1-score* 93%. Algoritma *Naive Bayes* dan *K-Nearest Neighbor* (KNN) memiliki performa yang lebih buruk yakni untuk *Naive Bayes* memiliki tingkat akurasi sebesar 72,75%, *precision* 69%, *recall* 96%, dan *f1-score* 80%, dan nilai akurasi sebesar 59,95%, *Precision* 94%, *Recall* 61%, dan *f1-score* 74% untuk algoritma knn. Dengan demikian pada penelitian ini, algoritma *Support Vector Machine*(SVM) dapat diandalkan untuk memprediksi keberlanjutan Gojek melalui analisis sentimen pengguna. Temuan ini memberikan wawasan berharga untuk pengembangan strategi bisnis Gojek di masa depan.

Kata kunci: Analisis Sentimen, Gojek, KNN, *Naive Bayes*, SVM

**PREDIKSI KEBERLANJUTAN GOJEK DI PASAR
TRANSPORTASI ONLINE INDONESIA MENGGUNAKAN
ANALISIS SENTIMEN DENGAN ALGORITMA NAÏVE
BAYES, KNN, DAN SVM**

ABSTRACT (English)

Online transportation has become a crucial need in the daily lives of Indonesians, with Gojek as the market leader. However, from 2021 to 2023, the number of Gojek application downloads has consistently declined. This raises concerns about Gojek's sustainability in the online transportation market. This study aims to predict Gojek's sustainability by analyzing user sentiment based on reviews from Google Play Store.

The method employed is sentiment analysis using three algorithms: Naive Bayes, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). The study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which includes data collection, data preprocessing, modeling, evaluation, and implementation stages. User reviews from Google Play Store are processed to classify sentiments as positive, negative, or neutral.

Research findings indicate that the Support Vector Machine (SVM) algorithm provides the best performance in sentiment classification with an accuracy rate of 89.65%, precision of 92%, recall of 93%, and an F1-score of 93%. The Naïve Bayes and K-Nearest Neighbor (KNN) algorithms performed worse, with Naïve Bayes achieving an accuracy rate of 72.75%, precision of 69%, recall of 96%, and an F1-score of 80%, while KNN obtained an accuracy rate of 59.95%, precision of 94%, recall of 61%, and an F1-score of 74%. Therefore, in this study, the Support Vector Machine (SVM) algorithm is deemed reliable for predicting Gojek's sustainability through user sentiment analysis. These findings provide valuable insights for developing Gojek's future business strategies.

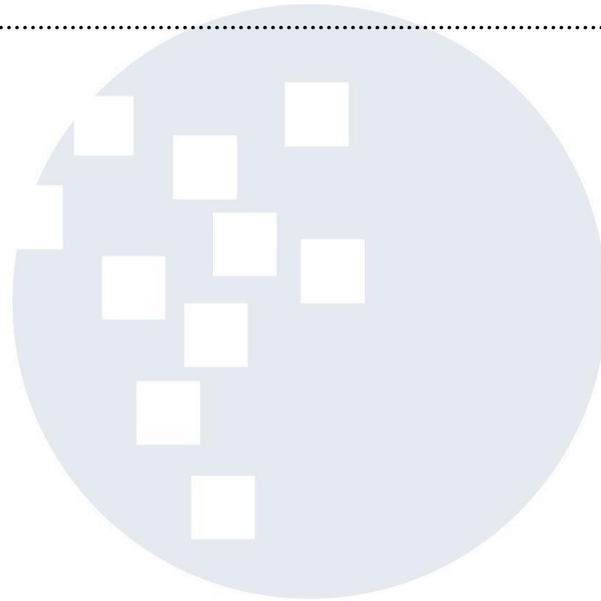
Keywords: *Sentiment Analysis, Gojek, KNN, Naïve Bayes, SVM*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian	6
1.4.2 Manfaat Penelitian	7
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Penelitian Terkait	9
2.2 Tinjauan Teori	16
2.2.1 Transportasi dan Transportasi Online	16
2.2.2 Analisis Sentimen	17
2.2.3 <i>Text Mining</i>	17
2.2.4 <i>Confusion Matrix</i>	18
2.2.5 Web Scraping	19
2.3 Framework dan Algoritma yang digunakan	19
2.3.1 CRISP-DM (Cross Industry Standard Process – Data Mining)	19
2.3.2 <i>Naïve Bayes</i>	21

2.3.3	<i>K-Nearest Neighbors (KNN)</i>	22
2.3.4	<i>Support Vector Machine (SVM)</i>	23
2.4	Tools yang digunakan	24
2.4.1	<i>Python</i>	24
BAB III METODOLOGI PENELITIAN		26
3.1	Gambaran Umum Objek Penelitian	26
3.2	Metode Penelitian	26
3.2.1	Alur Penelitian	26
3.2.2	Metode <i>Data Mining</i>	27
3.3	Teknik Pengumpulan Data	29
3.4	Variabel Penelitian	30
3.5	Teknik Analisis Data	30
BAB IV ANALISIS DAN HASIL PENELITIAN		33
4.1	<i>Business Understanding</i>	33
4.2	<i>Data Understanding</i>	33
4.3	<i>Data Preparation</i>	36
4.3.1	<i>Drop Duplicate Data</i>	37
4.3.2	<i>Cleaning Data</i>	38
4.3.3	<i>Preprocessing</i>	40
4.3.3	<i>Labeling</i>	45
4.3.3	<i>Over Sampling</i>	49
4.4	<i>Modeling</i>	51
4.4.1	<i>Word Cloud</i>	52
4.4.2	<i>Naïve Bayes</i>	54
4.4.2	<i>K-Nearest Neighbor (KNN)</i>	54
4.4.2	<i>Support Vector Machine (SVM)</i>	55
4.5	<i>Evaluation</i>	56
4.5.1	<i>Naive Bayes</i>	56
4.5.2	<i>K-Nearest Neighbor (KNN)</i>	60
4.5.3	<i>Support Vector Machine (SVM)</i>	63
4.6	<i>Deployment</i>	67

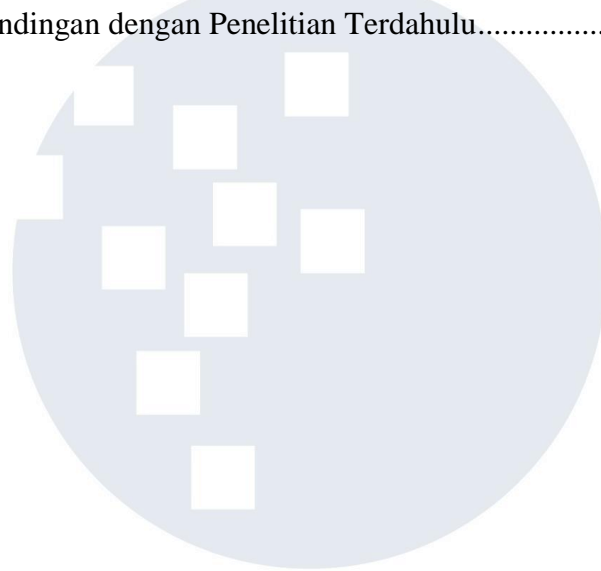
4.7 Diskusi dan Pembahasan	70
BAB V SIMPULAN DAN SARAN.....	73
5.1 Simpulan.....	73
5.2 Saran.....	74
DAFTAR PUSTAKA	76
LAMPIRAN.....	81



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	9
Tabel 3. 1 Perbandingan Algoritma Penelitian	28
Tabel 4. 1 Tabel Hasil Perhitungan Nilai Confusion Matrix dengan Naïve Bayes	58
Tabel 4. 2 Tabel Hasil Perhitungan Nilai Confusion Matrix dengan KNN.....	61
Tabel 4. 3 Tabel Hasil Perhitungan Nilai Confusion Matrix dengan SVM.....	65
Tabel 4. 4 Perbandingan dengan Penelitian Terdahulu.....	70



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

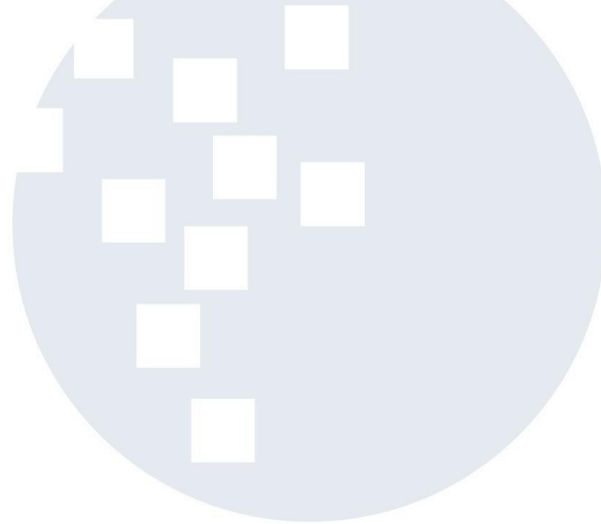
Gambar 4. 1 Potongan Kode Library yang digunakan.....	34
Gambar 4. 2 Potongan Kode Scraping Data yang Diperlukan.....	35
Gambar 4. 3 Data Hasil Scraping.....	35
Gambar 4. 4 Potongan Kode untuk Membaca Data.....	36
Gambar 4. 5 Informasi Mengenai Dataset	36
Gambar 4. 6 DataFrame yang Digunakan.....	37
Gambar 4. 7 Informasi DataFrame yang Digunakan	37
Gambar 4. 8 Menghapus Data Duplikat.....	38
Gambar 4. 9 Informasi DataFrame Setelah Menghilangkan Duplikat.....	38
Gambar 4. 10 Potongan Kode untuk Cleaning Data	39
Gambar 4. 11 Tabel Setelah Cleaning Data	40
Gambar 4. 12 Potongan Kode untuk Case Folding.....	40
Gambar 4. 13 Tabel setelah Case Folding	41
Gambar 4. 14 Potongan Kode untuk Mengubah Kata Tidak Baku.....	41
Gambar 4. 15 Data Kamus Kata Baku	42
Gambar 4. 16 Tabel Setelah Normalization	42
Gambar 4. 17 Mengatur DataFrame yang Diperlukan.....	43
Gambar 4. 18 Proses Tokenize.....	43
Gambar 4. 19 Import Library yang Digunakan Untuk Removing Stopwords	43
Gambar 4. 20 Proses Removing Stopwords.....	44
Gambar 4. 21 Import Library yang Digunakan Untuk Stemming Data.....	44
Gambar 4. 22 Proses Stemming Data	45
Gambar 4. 23 Potongan Kode untuk Menyimpan Data Hasil Preprocessing	45
Gambar 4. 24 Membaca Data Hasil Preprocessing.....	46
Gambar 4. 25 Validasi Data Hasil Preprocessing	46
Gambar 4. 26 Menghapus Data Kosong	47
Gambar 4. 27 Potongan Kode untuk Menentukan Sentimen Berdasarkan Kamus Leksikon yang ada.....	47
Gambar 4. 28 DataFrame Setelah Proses Labeling.....	48
Gambar 4. 29 Grafik Jumlah Sentimen.....	48
Gambar 4. 30 Potongan Kode untuk Menyimpan Data Hasil Labeling	49
Gambar 4. 31 DataFrame yang akan digunakan untuk Over Sampling.....	49
Gambar 4. 32 Jumlah Sentimen Sebelum Over Sampling	50
Gambar 4. 33 Potongan Kode Untuk Over Sampling.....	50
Gambar 4. 34 Jumlah Sentimen Sebelum dan Setelah Over Sampling	51
Gambar 4. 35 Potongan Kode untuk Menyimpan Data Hasil Oversampled	51
Gambar 4. 36 Word Cloud setelah Data Preparation.....	52
Gambar 4. 37 Potongan Data untuk Import Library yang Dibutuhkan.....	52
Gambar 4. 38 Splitting Data.....	53
Gambar 4. 39 Vectorization	53
Gambar 4. 40 Parameter Naïve Bayes	54

Gambar 4. 41 Parameter KNN	55
Gambar 4. 42 Parameter SVM	56
Gambar 4. 43 Evaluasi Naïve Bayes.....	57
Gambar 4. 44 Confusion Matrix Naïve Bayes	57
Gambar 4. 45 Evaluasi KNN	60
Gambar 4. 46 Confusion Matrix KNN.....	61
Gambar 4. 47 Evaluasi SVM	64
Gambar 4. 48 Confusion Matrix SVM.....	64
Gambar 4. 49 Tampilan Awal Aplikasi	67
Gambar 4. 50 Tampilan Data Preview	68
Gambar 4. 51 Tampilan Persebaran Sentimen.....	68
Gambar 4. 52 Tampilan Data Berdasarkan Filter Sentimen	69
Gambar 4. 53 Tampilan Data Berdasarkan Filter Rating Score.....	69



DAFTAR RUMUS

Rumus 2.1 Rumus Akurasi Pada Confusion Matrix [29]	18
Rumus 2.2 Rumus Precision Pada Confusion Matrix [29]	19
Rumus 2.3 Rumus Recall Pada Confusion Matrix [29].....	19
Rumus 2.4 Rumus F1-Score Pada Confusion Matrix	19
Rumus 2. 5 Rumus Teorema Bayes [32]	21
Rumus 2.6 Rumus Euclidean Disctance[33].....	23
Rumus 2.7 Rumus Kernel Linear SVM [25]	24



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	81
Lampiran B Form Konsultasi Bimbingan.....	90



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA