

ABSTRAK

Aplikasi transportasi online seperti Gojek sangat penting bagi masyarakat Indonesia. Ulasan pengguna di Play Store memberikan informasi tentang kepuasan dan pengalaman pengguna. Namun, banyaknya ulasan membuat analisis manual sulit dilakukan, sehingga dibutuhkan metode otomatis untuk analisis sentimen. Penelitian ini menggunakan Support Vector Machine (SVM) dengan model TF-IDF untuk menganalisis sentimen ulasan pengguna Gojek. Tujuannya adalah untuk mengklasifikasikan ulasan menjadi positif atau negatif dan menilai kinerja model TF-IDF dalam prediksi sentimen. Proses penelitian dimulai dengan mengumpulkan data ulasan melalui web scraping, lalu memprosesnya dengan pembersihan teks, penghapusan stopwords, dan tokenisasi. Model TF-IDF kemudian digunakan untuk memprediksi sentimen dengan dataset yang telah diproses. Dataset dibagi dalam rasio 80:20, 90:10, dan 70:30 untuk pelatihan dan pengujian. Model dievaluasi dengan metrik akurasi, precision, recall, dan F1-score. Dari 8000 ulasan pengguna Gojek yang dianalisis, model dengan kernel linear mencapai akurasi tertinggi 77,88% pada rasio 80:20 untuk hasil terbaik dari penelitian kali ini. Penelitian ini fokus pada analisis sentimen positif dan negatif terhadap aplikasi Gojek di Play Store.

Kata Kunci : Analisis sentimen, Gojek, Play Store, *Support Vector Machine (SVM)*, TF-IDF.

ABSTRACT

Online transportation applications like Gojek are essential for Indonesian society. User reviews on the Play Store provide information about user satisfaction and experience. However, the large volume of reviews makes manual analysis difficult, necessitating an automated method for sentiment analysis. This research uses Support Vector Machine (SVM) with a TF-IDF model to analyze sentiment in Gojek user reviews. The objective is to classify reviews as positive or negative and to evaluate the performance of the TF-IDF model in sentiment prediction. The research process begins with collecting review data through web scraping, followed by processing it through text cleaning, stopword removal, and tokenization. The TF-IDF model is then used to predict sentiment using the processed dataset. The dataset is divided into 80:20, 90:10, and 70:30 ratios for training and testing. The model is evaluated using accuracy, precision, recall, and F1-score metrics. From 8000 Gojek user reviews analyzed, the model with a linear kernel achieved the highest accuracy of 77.88% at an 80:20 ratio, representing the best outcome of this research. This study focuses on positive and negative sentiment analysis of the Gojek application on the Play Store.

Keywords: *Sentiment analysis, Gojek, Play Store, Support Vector Machine (SVM), TF-IDF.*

DAFTAR ISI

HALAMAN JUDUL	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
HALAMAN PERNYATAAN ORISINALITAS.....	ii
KATA PENGANTAR	iv
ABSTRAK.....	v
ABSTRACT	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR KODE PROGRAM.....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.2.1 Permasalahan	3
1.2.2 Metode Usulan.....	4
1.2.3 Pertanyaan Penelitian.....	4
1.3 Tujuan Dan Manfaat	4
1.3.1 Tujuan Penelitian	4
1.3.2 Manfaat Penelitian	5
1.4 Batasan-Batasan	5
1.5 Sintematika Penulisan Proposal	5
BAB II KAJIAN PUSTAKA	7
2.1 Landasan Teori	7
2.1.1 Natural Language Processing (NLP).....	7
2.1.2 Analisis Sentimen	7
2.1.3 Google Play Store	7
2.1.4 Gojek.....	8
2.1.5 Pengumpulan Data.....	8
2.1.6 Scraping Data.....	8
2.1.7 Preprocessing Data	9
2.1.8 TF-IDF (Term Frequency–Inverse Document Frequency).....	10
2.1.9 Support Vector Machine	10
2.1.10 Confusion matrix	12
2.2 Penelitian Terkait	13

BAB III METODE USULAN	15
3.1 Arsitektur Sistem	15
3.2 Data Penelitian	17
3.2.1 Dataset	17
3.2.2 Pelabelan Data	18
3.2.3 Preprocessing	20
3.2.4 Cleaning	20
3.2.5 Case folding	21
3.2.6 Tokenizing	21
3.2.7 Filtering.....	22
3.3 TF-IDF (Term Frequency-Inverse Document Frequency).....	22
3.4 Klasifikasi Support Vector Machine	27
3.4.1 Support Vector Machine dengan TF-IDF	29
3.5 Evaluasi.....	32
3.6 Skenario Pengujian	33
BAB IV HASIL DAN PEMBAHASAN	34
4.1 Pengumpulan Data	34
4.2 Pelabelan Data	38
4.3 Preprocessing Data.....	39
4.3.1 Cleaning.....	39
4.3.2 Case Folding.....	40
4.3.3 Tokenizing	41
4.3.4 Filtering	42
4.4 Ekstaksi Fitur	44
4.5 Pembagian Data	46
4.6 Support Vectore Machine.....	47
4.7 Confusion Matrix	48
4.8 Hasil Skenario Uji Coba	49
4.9 Implementasi Streamlit	53
4.10 Analisis Hasil Uji Coba	55
BAB V PENUTUP	57
5.1 Kesimpulan	57
5.2 Saran	57
DAFTAR PUSTAKA	58

DAFTAR GAMBAR

Gambar 3.1 Arsitektur	15
Gambar 3.2 WordCloud	18
Gambar 3.3 Grafik Jumlah Ulasan Negatif	18
Gambar 3.4 Alur Klasifikasi SVM	27
Gambar 3.5 Inisialisasi Nilai Parameter	30
Gambar 4.1 Tampilan Awal Google Colab	34
Gambar 4.2 Tampilan Awal Lebar Kerja	34
Gambar 4.3 Dataframe.....	37
Gambar 4.5 Hasil Cleaning.....	40
Gambar 4.6 Hasil Case Folding.....	41
Gambar 4.7 Hasil Tokenizing	42
Gambar 4.8 Stopword.....	44
Gambar 4.9 Ekstraksi Fitur TF-IDF	45
Gambar 4.10 Pembagian Data	47
Gambar 4.11 Implementasi 1	54
Gambar 4.13 Akurasi Perbandingan Model	55

DAFTAR TABEL

Tabel 2.1 Model Confusion Matrix	12
Tabel 2.2 Penelitian Terkait	13
Tabel 3.1 Contoh Data Ulasan	17
Tabel 3.2 Pelabelan Data	19
Tabel 3.3 Cleaning	20
Tabel 3.4 Case Folding	21
Tabel 3.6 Filtering.....	22
Tabel 3.7 Hasil Preprocessing.....	23
Tabel 3.8 Data Hasil IDF	23
Tabel 3.9 Data Hasil TF-IDF	24
Tabel 3.10 TF-IDF dengan SVM.....	29
Tabel 3.11 Menghitung nilai kernel	30
Tabel 3.12 Menghitung Matriks	31
Tabel 3.13 Mencari nilai error	31
Tabel 3.14 Menghitung nilai delta alpha	31
Tabel 3.15 Mencari nilai alpha baru	31
Tabel 3.16 Mencari nilai bias.....	31
Tabel 3.17 Hasil perhitungan bias	32
Tabel 3.18 Menghitung nilai kernel data	32
Tabel 3.19 Pengujian	32
Tabel 3.20 Skenario Pengujian	33
Tabel 4.1 Sampel Data	38
Tabel 4.2 Hasil Uji Coba 1	49
Tabel 4.3 Hasil Uji Coba 2	50
Tabel 4.4 Hasil Uji Coba 3	51
Tabel 4.5 Hasil Uji Coba 4	51
Tabel 4.6 Hasil Uji Coba 5	52
Tabel 4.7 Hasil Uji Coba 6	53
Tabel 4.8 Keseluruhan Hasil Uji Coba	53

DAFTAR KODE PROGRAM

Kode Program 4.1 Scrapping data	35
Kode Program 4.2 Dataframe	36
Kode Program 4.3 Simpan Data	37
Kode Program 4.4 Cleaning	39
Kode Program 4.5 Case Folding.....	40
Kode Program 4.6 Tokenizing	41
Kode Program 4.7 Remove Stopword	43
Kode Program 4.8 Ekstraksi Fitur TF-IDF	44
Kode Program 4.9 Pembagian Data.....	46
Kode Program 4.10 SVM	47
Kode Program 4.1 Confusion Matriks.....	48