

**ANALISIS SENTIMEN APLIKASI DUOLINGO PADA GOOGLE
PLAY MENGGUNAKAN MODEL KLASIFIKASI**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

SALASIA FATMA SYFANA

21.12.2126

Kepada

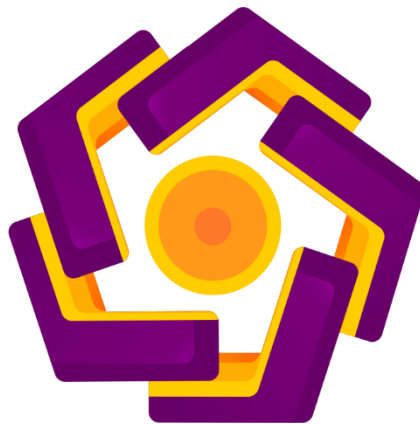
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN APLIKASI DUOLINGO PADA GOOGLE
PLAY MENGGUNAKAN MODEL KLASIFIKASI**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

SALASIA FATMA SYFANA

21.12.2126

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN APLIKASI DUOLINGO PADA GOOGLE PLAY
MENGUNAKAN MODEL KLASIFIKASI**

yang disusun dan diajukan oleh

Salasia Fatma Syfana

21.12.2126

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 17 Februari 2025

Dosen Pembimbing,



Yoga Pristvanto, S.Kom., M.Eng
NIK. 190302412

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN APLIKASI DUOLINGO PADA GOOGLE PLAY
MENGGUNAKAN MODEL KLASIFIKASI

yang disusun dan diajukan oleh

Salasia Fatma Syfana

21.12.2126

Telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Agung Nugroho, S.Kom., M.Kom.
NIK. 190302242

Ikamah, S.Kom., M.Kom
NIK. 190302282

Yoga Pristyanto, S.Kom., M.Eng.
NIK. 190302412



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Salasia Fatma Syfana

NIM : 21.12.2126

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Aplikasi Duolingo Pada Google Play Menggunakan Model Klasifikasi

Dosen Pembimbing : Yoga Pristyanto, S.Kom., M.Eng

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, 17 Februari 2025

Yang Menyatakan,

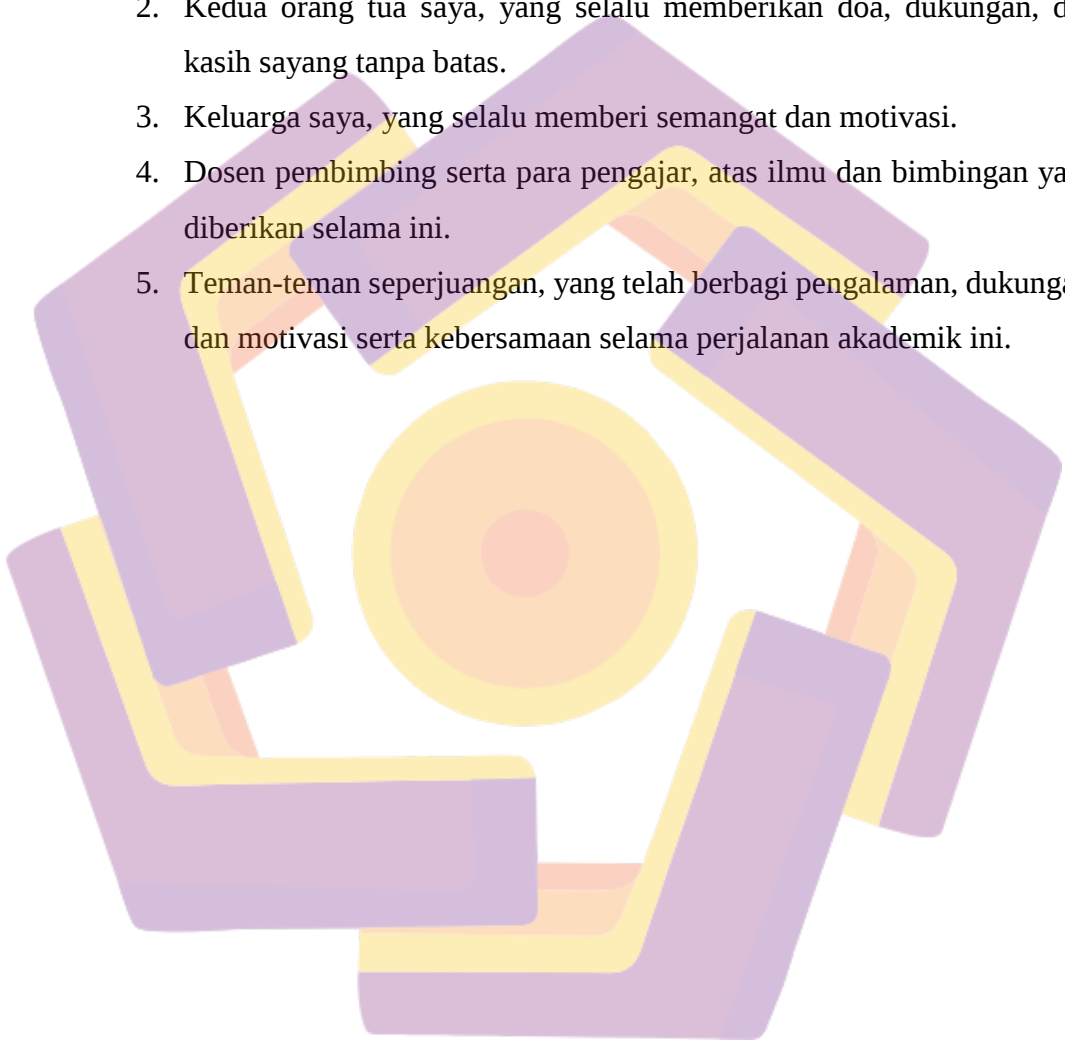


Salasia Fatma Syfana

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan kepada:

1. Allah SWT, atas segala rahmat, hidayah, dan kemudahan yang diberikan dalam setiap langkah kehidupan ini.
2. Kedua orang tua saya, yang selalu memberikan doa, dukungan, dan kasih sayang tanpa batas.
3. Keluarga saya, yang selalu memberi semangat dan motivasi.
4. Dosen pembimbing serta para pengajar, atas ilmu dan bimbingan yang diberikan selama ini.
5. Teman-teman seperjuangan, yang telah berbagi pengalaman, dukungan, dan motivasi serta kebersamaan selama perjalanan akademik ini.



KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “ dengan baik, lancar, dan tepat waktu. Skripsi ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Program Fakultas Ilmu Komputer
3. Bapak Anggit Dwi Hartanto, M.Kom selaku Ketua Program Studi Sistem Informasi
4. Bapak Yoga Pristyanto, S.Kom, M.Eng selaku dosen pembimbing yang memberikan arahan, saran, dan motivasi terhadap penulis
5. Kedua orang tua, keluarga besar, dan teman-teman tercinta yang memberikan semangat dan doa kepada penulis.
6. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan dukungan, bantuan, serta motivasi dalam penyelesaian skripsi ini.

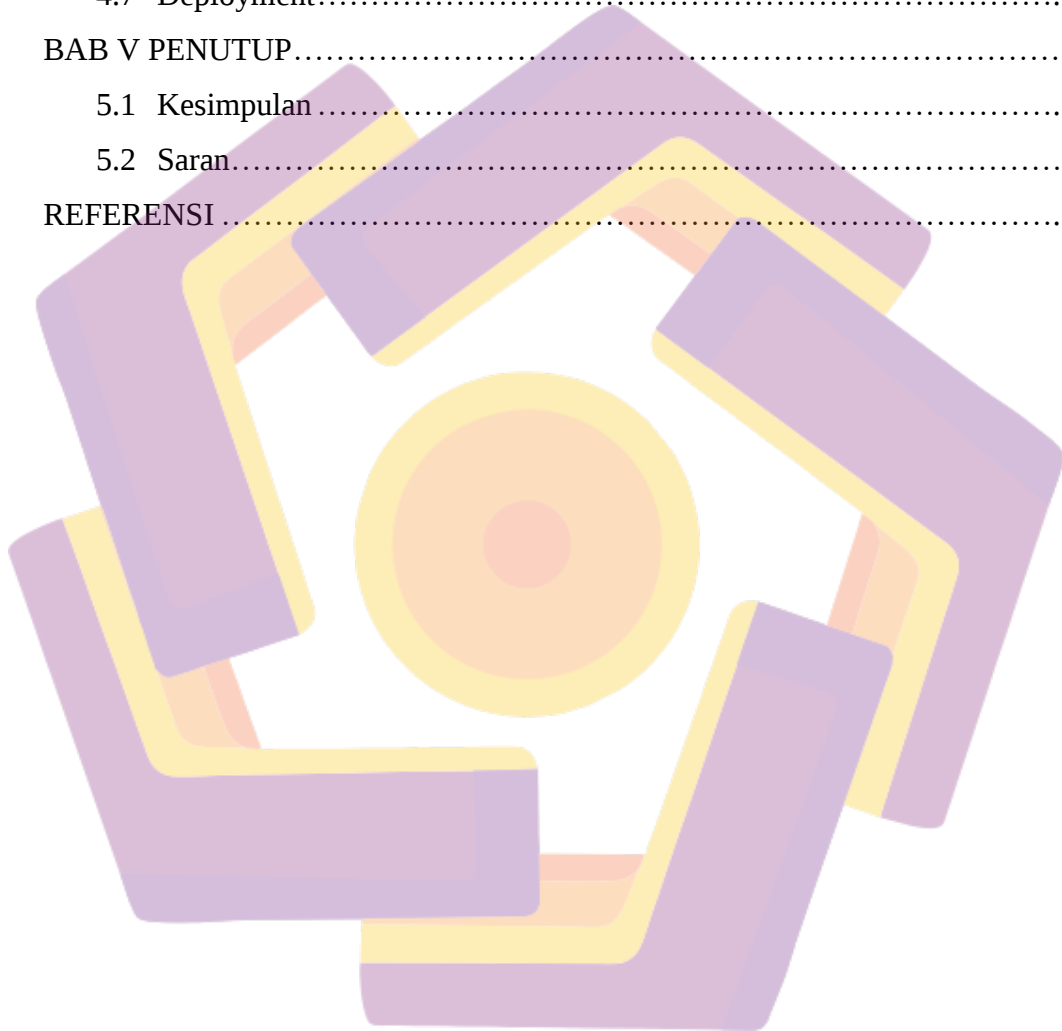
Yogyakarta, 17 Februari 2025

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.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH.....	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5.. Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	14
BAB III METODE PENELITIAN	25
3.1 Alur Penelitian	25
3.2 Alat dan Bahan.....	35
BAB IV HASIL DAN PEMBAHASAN	37
4.1 Scraping Data.....	37

4.2	Pra Pemrosesan	38
4.3	Labeling Data	48
4.4.	Split Data	50
4.5	TF-IDF	50
4.6	Klasifikasi dan Confusion Matrix	51
4.7	Deployment	57
BAB V PENUTUP		59
5.1	Kesimpulan	59
5.2	Saran	59
REFERENSI		60

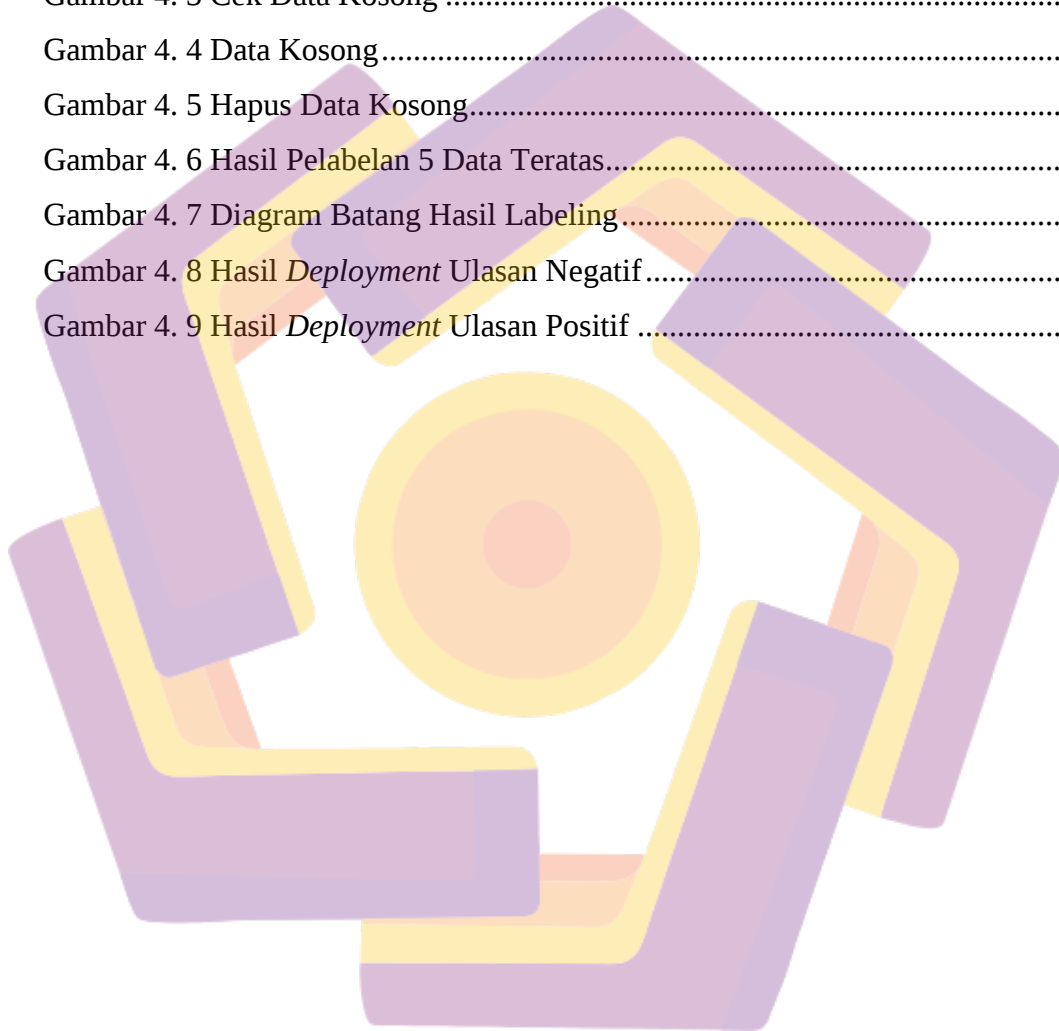


DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	8
Tabel 2. 2 <i>Confusion Matrix</i>	22
Tabel 3. 1 Contoh Tahap <i>Cleaning</i>	26
Tabel 3. 2 Contoh Tahap <i>Case Folding</i>	27
Tabel 3. 3 Contoh Tahap <i>Tokenizing</i>	27
Tabel 3. 4 Contoh Tahap <i>Stopword Removal</i>	28
Tabel 3. 5 Contoh Tahap <i>Stemming</i>	29
Tabel 3. 6 Contoh Perhitungan TF	30
Tabel 3. 7 Contoh Perhitungan IDF	32
Tabel 3. 8 Contoh Hasil Perhitungan TF-IDF.....	33
Tabel 4. 1 Hasil <i>Cleaning</i>	39
Tabel 4. 2 Hasil <i>Cleaning</i>	41
Tabel 4. 3 Hasil <i>Tokenizing</i>	42
Tabel 4. 4 Hasil <i>Stopword Removal</i>	44
Tabel 4. 5 Hasil <i>Stemming</i>	46
Tabel 4. 6 Contoh Hasil TF-IDF	50
Tabel 4. 7 Confusion Matrix Naïve Bayes.....	51
Tabel 4. 8 Confusion Matrix K-Nearest Neighbors.....	52
Tabel 4. 9 Confusion Matrix Support Vector Machine.....	54
Tabel 4. 10 Confusion Matrix Random Forest.....	55
Tabel 4. 11 Hasil Evaluasi Model Klasifikasi.....	56

DAFTAR GAMBAR

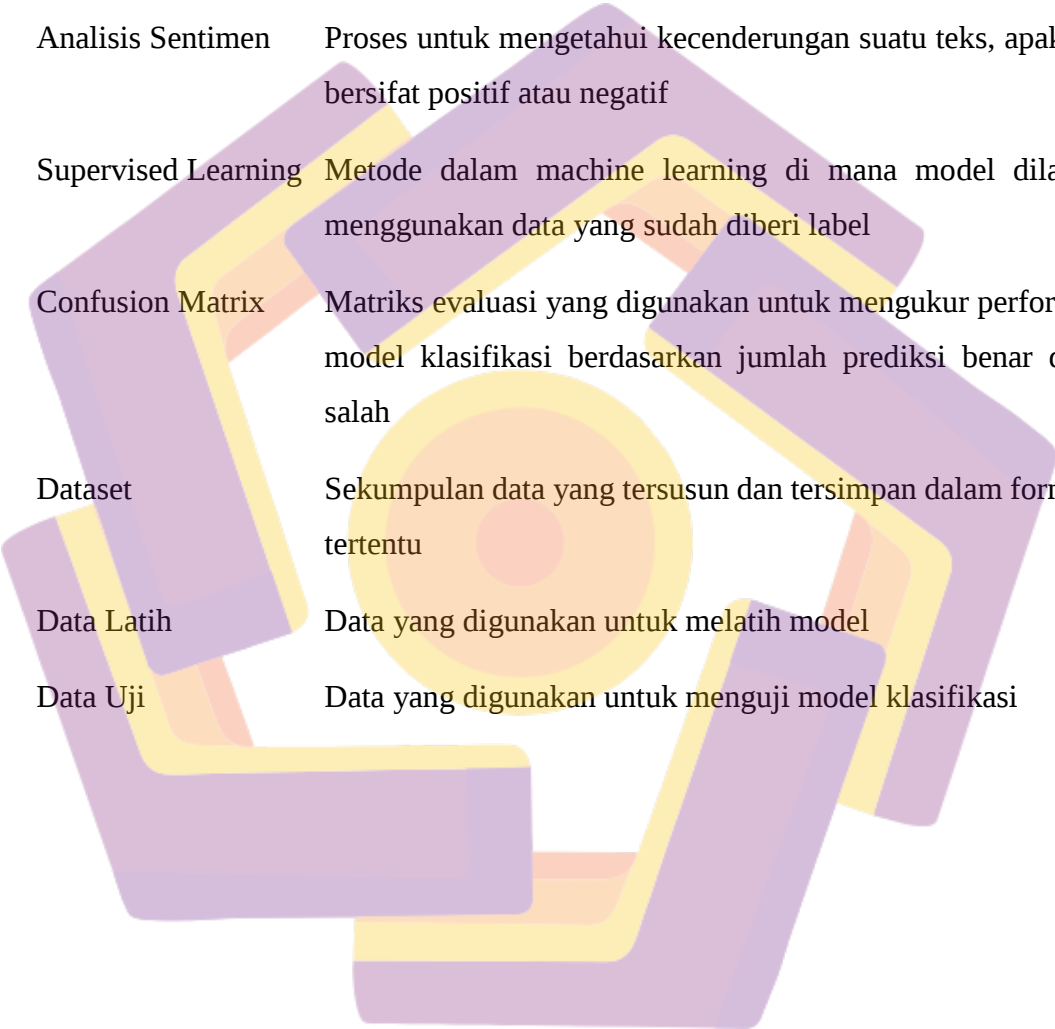
Gambar 3. 1 Alur Penelitian.....	25
Gambar 4. 1 Hasil Scraping Data Ulasan Aplikasi Duolingo.....	37
Gambar 4. 2 Dataset Ulasan dan Skor	38
Gambar 4. 3 Cek Data Kosong	47
Gambar 4. 4 Data Kosong.....	48
Gambar 4. 5 Hapus Data Kosong.....	48
Gambar 4. 6 Hasil Pelabelan 5 Data Teratas.....	49
Gambar 4. 7 Diagram Batang Hasil Labeling.....	49
Gambar 4. 8 Hasil <i>Deployment</i> Ulasan Negatif.....	57
Gambar 4. 9 Hasil <i>Deployment</i> Ulasan Positif	58



DAFTAR LAMBANG DAN SINGKATAN

TF-IDF	Term Frequency - Inverse Document Frequency
SVM	Support Vector Machine
KNN	K-Nearest Neighbor
ML	Machine Learning
NLP	Natural Language Processing
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
HTML	HyperText Markup Language
CSV	Comma Separated Values

DAFTAR ISTILAH



Klasifikasi	Pengelompokkan data ke dalam kategori tertentu
Model Klasifikasi	Algoritma dalam machine learning yang digunakan untuk mengelompokkan data ke dalam kategori tertentu
Analisis Sentimen	Proses untuk mengetahui kecenderungan suatu teks, apakah bersifat positif atau negatif
Supervised Learning	Metode dalam machine learning di mana model dilatih menggunakan data yang sudah diberi label
Confusion Matrix	Matriks evaluasi yang digunakan untuk mengukur performa model klasifikasi berdasarkan jumlah prediksi benar dan salah
Dataset	Sekumpulan data yang tersusun dan tersimpan dalam format tertentu
Data Latih	Data yang digunakan untuk melatih model
Data Uji	Data yang digunakan untuk menguji model klasifikasi

INTISARI

Banyaknya negara di belahan dunia ini menyebabkan munculnya berbagai macam bahasa yang tersebar di seluruh penjuru dunia. Keanekaragaman bahasa ini seringkali menjadi hambatan bagi banyak orang untuk berkomunikasi dengan individu dari negara lain, terutama jika tidak menguasai bahasa asing Negara tersebut. Namun, dengan kemajuan teknologi saat ini, banyak media elektronik yang dikembangkan untuk membantu pembelajaran bahasa asing. Salah satu aplikasi populer untuk pembelajaran bahasa asing adalah aplikasi DuoLingo, yang dirancang untuk memudahkan pengguna dalam mempelajari berbagai bahasa. Keberhasilan suatu aplikasi pembelajaran bahasa dapat diukur dengan melakukan analisis sentimen terhadap komentar dan ulasan dari para pengguna untuk aplikasi tersebut. Pada penelitian ini, ulasan yang akan dianalisis diambil dari ulasan terhadap aplikasi Duolingo pada Google Play Store sebanyak 1999 ulasan, setelah dilakukan preprocessing, dengan sentimen positif sebanyak 1696 dan sentimen negatif sebanyak 303 ulasan. Pada penelitian ini empat model klasifikasi yaitu K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machine (SVM) dan Random Forest akan dibandingkan untuk mengetahui model yang paling baik. Hasil penelitian menunjukkan bahwa model yang paling baik dalam mengklasifikasikan ulasan positif dan negatif terhadap aplikasi Duolingo adalah model Support Vector Machine dengan akurasi sebesar 88.5%.

Kata kunci : Analisis Sentimen, Duolingo, Google Play Store, Model Klasifikasi

ABSTRACT

The large number of countries in this part of the world has led to the emergence of various languages spread throughout the world. This linguistic diversity often becomes an obstacle for many people to communicate with individuals from other countries, especially if they do not master the foreign language of that country. However, with current technological advances, many electronic media have been developed to assist foreign language learning. One of the popular applications for learning foreign languages is the Duolingo application, which is designed to make it easier for users to learn various languages. The success of a language learning application can be measured by conducting sentiment analysis of comments and reviews from users for the application. In this research, the reviews that will be analyzed are taken from reviews of the Duolingo application on the Google Play Store totaling 1999 reviews, after preprocessing, with 1696 positive sentiments and 303 negative sentiments. In this research, four classification models, namely K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machine (SVM) and Random Forest, will be compared to find out the best model. The research results show that the best model for classifying positive and negative reviews of the Duolingo application is the Support Vector Machine model with an accuracy of 88.5%.

Keywords: Sentiment Analysis, Duolingo, Google Play Store, Classification Model