

**KLASIFIKASI ULASAN PRODUK PADA PESAING TOKO
DEPESAFLOREST MENGGUNAKAN ALGORITMA
NAÏVE BAYES**

SKRIPSI

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



disusun oleh

TATAK GORESTA PUTRA NANDA

18.11.2442

Kepada

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

**KLASIFIKASI ULASAN PRODUK PADA PESAING TOKO
DEPESAFLOREST MENGGUNAKAN ALGORITMA
NAÏVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

TATAK GORESTA PUTRA NANDA

18.11.2442

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI ULASAN PRODUK PADA PESAING TOKO
DEPESAFLOREST MENGGUNAKAN ALGORITMA NAÏVE BAYES**

yang disusun dan diajukan oleh

Tatak Goresta Putra Nanda

18.11.2442

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 07 Mei 2025

Dosen Pembimbing,



Erni Seniwati, S.Kom, M.Cs
NIK. 190302231

HALAMAN PENGESAHAN
SKRIPSI
KLASIFIKASI ULASAN PRODUK PADA PESAING TOKO
DEPESAFLOREST MENGGUNAKAN ALGORITMA NAÏVE BAYES

yang disusun dan diajukan oleh

Tatak Goresta Putra Nanda

18.11.2442

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Mei 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Subektiningsih, S.Kom., M.Kom.
NIK. 190302413



Uyock Anggoro Saputro, S.Kom., M.Kom.
NIK. 190302419



Eli Pujastuti, S.Kom., M.Kom.
NIK. 190302227



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 Mei 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Tatak Goresta Putra Nanda
NIM : 18.11.2442

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Ulasan Produk Pada Pesaing Toko Depesaflorist Menggunakan Algoritma Naïve Bayes

Dosen Pembimbing : Erni Seniwati, S.Kom, M.Cs

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, Jumat, 23 Mei 2025

Yang Menyatakan,



Tatak Goresta Putra Nanda

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam atas rahmat dan karunia Allah SWT yang telah memberi kemudahan, kesehatan dan kelancaran kepada peneliti dalam menyelesaikan skripsi ini dengan sebaik-baiknya. Karya ini menjadi persembahan bagi orang-orang yang peneliti kasihi dan sayangi.

1. Kepada orang tua saya, Bapak Dalimin dan Ibu Sukapjiah, yang sudah mau menunggu saya sampai saya menyelesaikan skripsi ini. Terima kasih atas motivasi, dukungan, dan doa yang tidak pernah putus. Semoga Allah SWT selalu memberi kalian kesehatan dan umur yang panjang untuk menemaniku hingga menuju kesuksesan.
2. Kepada seluruh keluarga yang tidak dapat saya sebutkan satu per satu, terima kasih saya ucapkan atas segala kebaikan dan dukungan yang telah diberikan kepada saya.
3. Kepada Ibu Erni Seniwati, S.Kom, M.Cs yang sudah membimbing saya selama penelitian ini berlangsung.
4. Kepada teman baik saya selama masa perkuliahan, Gilang Nur Fitriyanto, Hanif Ardhiansyah Yudith, Septa Najibul Wahid, Dimas Pramana Indratama, Ardiyanto Saputra, Teguh Efriyanto, Riko Aditya Nurtito, Gilang Anugrah Multi, Riko Adi Septiawan. Terima kasih atas kebersamaan, semangat, dan dukungan yang telah kalian berikan
5. Kepada seluruh teman 18 Informatika 09, yang telah berjuang berbagi ilmu dan banyak hal.
6. Serta semua pihak yang tidak dapat disebutkan satu-satu.

KATA PENGANTAR

Puji dan syukur bagi Allah SWT atas berkah, rahmat, dan nikmat yang diberikan kepada hamba-Nya sehingga peneliti dapat menyelesaikan penyusunan tugas akhir yang berjudul “Klasifikasi Ulasan Produk Pada Pesaing Toko Depesaflorist Menggunakan Algoritma Naïve Bayes”. Skripsi ini disusun sebagai syarat mendapatkan gelar Sarjana Komputer dari Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Tidak dapat disangkal bahwa butuh usaha yang keras dalam penyelesaian penyusunan skripsi ini. Namun, karya ini tidak akan selesai tanpa orang-orang tercinta di sekeliling peneliti yang mendukung dan membantu. Terimakasih peneliti sampaikan kepada:

1. Tuhan yang Maha Esa, Allah SWT kerana atas rahmat, hidayah, serta karunia-Nya yang telah diberikan kepada penulis sehingga tugas akhir ini dapat terselesaikan.
2. Bapak Prof. Dr. M. Suyanto, MM. Selaku Rektor Universitas Amikom Yogyakarta.
3. Ibu Erni Seniwati, S.Kom, M.Cs selaku dosen pembimbing yang telah memberikan banyak arahan bagi peneliti dalam penyusunan penelitian ini.
4. Kedua orang tua penulis dan keluarga yang telah mendukung dan selalu mendo'akan penulis selama ini.
5. Para Dosen dan Staff Universitas Amikom Yogyakarta yang telah berperan dalam memberikan ilmu pengetahuan dan pengalaman selama masa perkuliahan.
6. Semua pihak yang telah membantu dalam penyelesaian skripsi ini yang tidak bisa disebutkan satu per satu oleh penulis.

Yogyakarta, 23 Mei 2025

Penulis

DAFTAR ISI

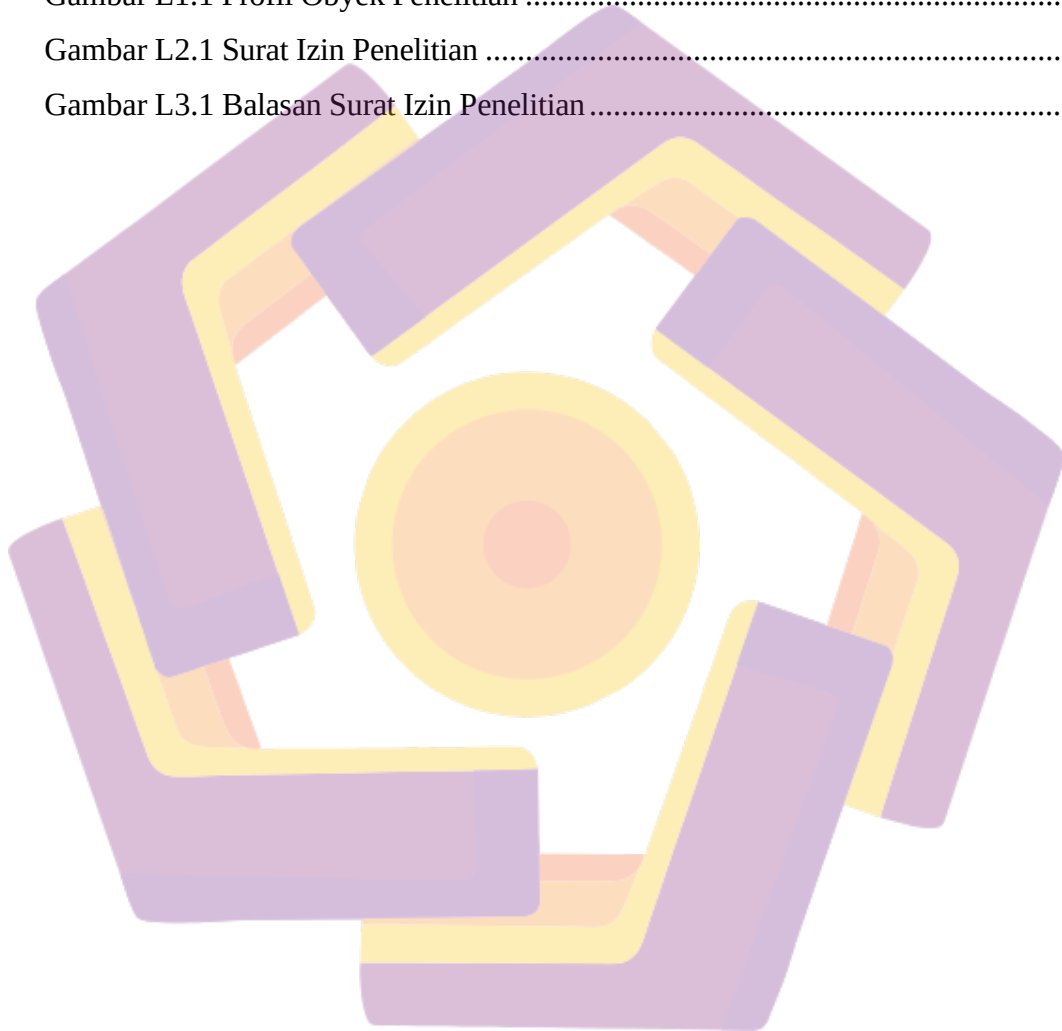
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xv
DAFTAR LAMBANG DAN SINGKATAN	xvi
DAFTAR ISTILAH.....	xvii
INTISARI	xviii
ABSTRACT.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Literatur Review	6
Tabel 2.1 Perbandingan Penelitian.....	8
2.2 Landasan Teori	10
2.2.1 Data Mining.....	10
2.2.2 Analisis Sentimen	10
2.2.3 Text Mining	11
2.2.4 Preprocessing.....	11
2.2.5 Term Frequency-Inverse Document Frequency (TF-IDF)	12
2.2.6 Naïve Bayes	13
2.2.7 Confussion Matrix	14
Tabel 2.2 Confusion Matrix	14
2.2.8 Flowchart	15

Tabel 2.3 Daftar Simbol Dokumen Flowchart	15
2.2.9 Python	16
2.3.0 Shopee API	16
BAB III METODE PENELITIAN	18
3.1 Objek Penelitian	18
3.2 Alat dan Bahan	18
3.2.1 Bahan Penelitian	18
3.2.1 Alat Penelitian	18
3.3 Alur Penelitian	19
Gambar 3.1 Flowchart Alur Penelitian	19
3.4 Langkah-Langkah Penelitian	19
3.4.1 Scrapping Data Shopee	19
Gambar 3.2 Diagram Alir Scrapping Data	20
3.4.2 Pelabelan Data	20
3.4.3 Preprocessing	20
Gambar 3.3 Diagram Alir Preprocessing	20
3.4.4 Pembagian Data	23
3.4.5 Pembobotan TF-IDF	23
3.4.6 Model Klasifikasi Naïve Bayes	23
Gambar 3.4 Diagram Alir Klasifikasi	24
3.5 Analisis Algoritma	25
3.5.1 Pengumpulan Data	25
Tabel 3.1 Data tweet yang telah diperoleh	25
3.5.2 Pelabelan Data	26
Tabel 3.2 Data ulasan yang telah diberi label	26
3.5.3 Cleaning Data	26
Tabel 3.3 Cleaning Data	27
3.5.4 Case Folding	28
Tabel 3.4 Case Folding	28
3.5.5 Normalisasi Kata	29
Tabel 3.5 Normalisasi Kata	29
3.5.6 Tokenizing	30
Tabel 3.6 Tokenizing	30
3.5.7 Stopword Removal	31
Tabel 3.7 Stopword Removal	32
3.5.8 Stemming	33
Tabel 3.8 Stemming	33
3.5.9 Pembobotan Term Frequency-Inverse Document Frequency	34
Tabel 3.9 Term Hasil Preprocessing	34

Tabel 4.0 Term Frequency Kelas Positif	35
Tabel 4.1 Term Frequency Kelas Negatif.....	36
Tabel 4.2 Term Frequency Kelas Netral.....	37
Tabel 4.3 Document Frequency	39
Tabel 4.4 Inverse Document Frequency	42
Tabel 4.5 Term Frequency-Inverse Document Frequency Positif.....	46
Tabel 4.6 Term Frequency-Inverse Document Frequency Negatif	47
Tabel 4.7 Term Frequency-Inverse Document Frequency Netral	49
Tabel 4.8 Hasil Akhir TF-IDF	51
3.5.10 Klasifikasi Naïve Bayes	52
3.5.11 Contoh Perhitungan Naïve Bayes.....	53
3.5.12 Evaluasi dengan Confusion Matrix	57
Tabel 4.9 Confusion Matrix	57
BAB IV HASIL DAN PEMBAHASAN	58
4.1 Pengambilan Data Ulasan.....	58
Gambar 4.1 Script Scrapping	59
Gambar 4.2 Sumber Dataset	59
Gambar 4.3 Hasil Scrapping	60
4.1.1 API Shopee	60
Gambar 4.4 Tampilan API Shopee	61
4.2 Pelabelan Data	61
Gambar 4.5 Pelabelan Data	62
4.3 Preprocessing	62
4.3.1 Cleaning.....	63
Gambar 4.6 Script Cleaning	63
Gambar 4.7 Hasil Cleaning.....	64
4.3.2 Case Folding	64
Gambar 4.8 Script Case Folding	64
Gambar 4.9 Hasil Script Case Folding.....	65
4.3.3 Normalisasi Kata	65
Gambar 4.10 Contoh Kamus Normalisasi	66
Gambar 4.11 Script Normalisasi.....	66
Gambar 4.12 Hasil Normalisasi	66
4.3.4 Tokenizing	66
Gambar 4.13 Script Tokenizing	67

Gambar 4.14 Hasil Tokenizing	67
4.3.5 Stopword Removal	67
Gambar 4.15 Script Stopword Removal	68
Gambar 4.16 Hasil Stopword Removal	69
4.3.6 Stemming	69
Gambar 4.17 Script Stemming	70
Gambar 4.18 Hasil Stemming	70
4.3.7 Hasil Akhir Preprocessing	70
Gambar 4.19 Script Hasil Akhir Preprocessing	71
Gambar 4.20 Hasil Akhir Preprocessing	71
4.4 Pembagian Data	71
Gambar 4.21 Script Pembagian Data	72
Gambar 4.22 Hasil Pembagian Data Training	72
Gambar 4.23 Hasil Pembagian Data Testing	73
4.5 Pembobotan Term Frequency-Inverse Document Frequency	73
Gambar 4.24 Script TF-IDF	73
Gambar 4.25 Hasil Perhitungan TF-IDF Training	74
Gambar 4.26 Hasil Perhitungan TF-IDF Testing	74
Gambar 4.27 Script TF-IDF Berdasarkan Kelas Sentimen	74
Gambar 4.28 Hasil Perhitungan TF-IDF Berdasarkan Kelas Sentimen	75
4.6 Model Klasifikasi Naïve Bayes	75
Gambar 4.29 Script Model Klasifikasi	76
4.7 Evaluasi Model Klasifikasi Naïve Bayes	76
Gambar 4.30 Script Evaluasi Model Klasifikasi	77
Gambar 4.31 Hasil Evaluasi Model Klasifikasi	77
4.8 Confusion Matrix	77
Gambar 4.32 Script Confusion Matrix	78
Gambar 4.33 Hasil Confusion Matrix	79
4.9 Visualisasi Data	79
Gambar 4.34 Script Wordcloud	80
Gambar 4.35 Hasil Wordcloud Positif	80
Gambar 4.36 Hasil Wordcloud Negatif	81
Gambar 4.37 Hasil Wordcloud Netral	81
Gambar 4.38 Script Pie Chart	82

Gambar 4.39 Hasil Pie Chart.....	82
BAB V PENUTUP	83
5.1 Kesimpulan	83
5.2 Saran	84
DAFTAR PUSTAKA	86
LAMPIRAN.....	89
Gambar L1.1 Profil Obyek Penelitian	89
Gambar L2.1 Surat Izin Penelitian	90
Gambar L3.1 Balasan Surat Izin Penelitian.....	91



DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian.....	8
Tabel 2.2 Confusion Matrix	14
Tabel 2.3 Daftar Simbol Dokumen Flowchart	15
Tabel 3.1 Data tweet yang telah diperoleh.....	25
Tabel 3.2 Data ulasan yang telah diberi label.....	26
Tabel 3.3 Cleaning Data.....	27
Tabel 3.4 Case Folding	28
Tabel 3.5 Normalisasi Kata.....	29
Tabel 3.6 Tokenizing	30
Tabel 3.7 Stopword Removal.....	32
Tabel 3.8 Stemming	33
Tabel 3.9 Term Hasil Preprocessing.....	34
Tabel 4.0 Term Frequency Kelas Positif	35
Tabel 4.1 Term Frequency Kelas Negatif.....	36
Tabel 4.2 Term Frequency Kelas Netral.....	37
Tabel 4.3 Document Frequency	39
Tabel 4.4 Inverse Document Frequency	42
Tabel 4.5 Term Frequency-Inverse Document Frequency Positif.....	46
Tabel 4.6 Term Frequency-Inverse Document Frequency Negatif.....	47
Tabel 4.7 Term Frequency-Inverse Document Frequency Netral	49
Tabel 4.8 Hasil Akhir TF-IDF	51
Tabel 4.9 Confusion Matrix	57

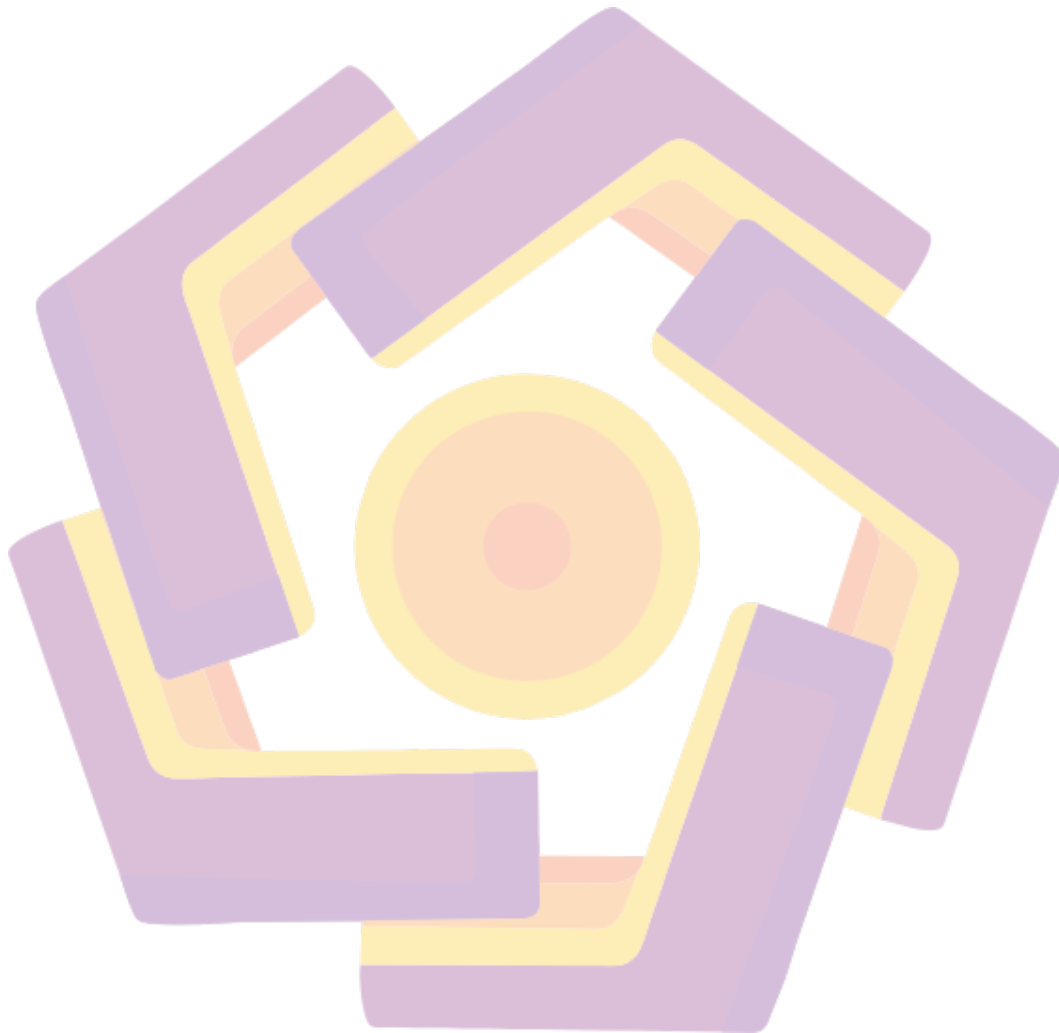
DAFTAR GAMBAR

Gambar 3.1 Flowchart Alur Penelitian	19
Gambar 3.2 Diagram Alir Scrapping Data.....	20
Gambar 3.3 Diagram Alir Preprocessing	20
Gambar 3.4 Diagram Alir Klasifikasi	24
Gambar 4.1 Script Scrapping	59
Gambar 4.2 Sumber Dataset	59
Gambar 4.3 Hasil Scrapping	60
Gambar 4.4 Tampilan Api Shopee.....	61
Gambar 4.5 Pelabelan Data.....	62
Gambar 4.6 Script Cleaning.....	63
Gambar 4.7 Hasil Cleaning.....	64
Gambar 4.8 Script Case Folding	64
Gambar 4.9 Hasil Case Folding	65
Gambar 4.10 Contoh Kamus Normalisasi	66
Gambar 4.11 Script Normalisasi	66
Gambar 4.12 Hasil Normalisasi	66
Gambar 4.13 Script Tokenizing	67
Gambar 4.14 Hasil Tokenizing	67
Gambar 4.15 Script Stopword Removal	68
Gambar 4.16 Hasil Stopword Removal	69
Gambar 4.17 Script Stemming.....	70
Gambar 4.18 Hasil Stemming.....	70
Gambar 4.19 Script Hasil Akhir Preprocessing	71
Gambar 4.20 Hasil Akhir Preprocessing.....	71
Gambar 4.21 Script Pembagian Data	72
Gambar 4.22 Hasil Pembagian Data Training	72
Gambar 4.23 Hasil Pembagian Data Testing	73
Gambar 4.24 Script TF-IDF	73
Gambar 4.25 Hasil Perhitungan TF-IDF Training.....	74

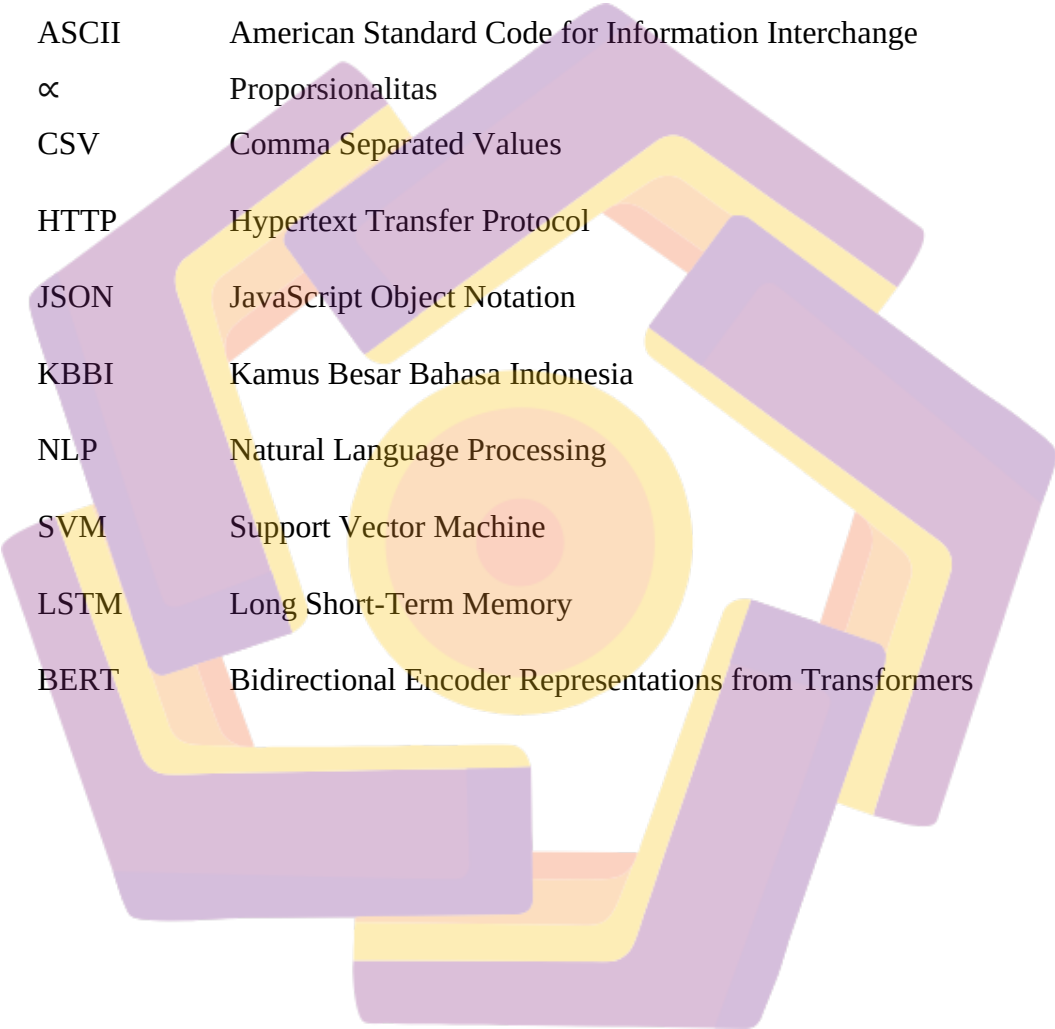
Gambar 4.26 Hasil Perhitungan TF-IDF Testing.....	74
Gambar 4.27 Script TF-IDF Berdasarkan Kelas Sentimen.....	74
Gambar 4.28 Hasil Perhitungan TF-IDF Berdasarkan Kelas Sentimen	75
Gambar 4.29 Script Model Klasifikasi	76
Gambar 4.30 Script Evaluasi Model Klasifikasi.....	77
Gambar 4.31 Hasil Evaluasi Model Klasifikasi.....	77
Gambar 4.32 Script Confusion Matrix.....	78
Gambar 4.33 Hasil Confusion Matrix.....	79
Gambar 4.34 Script Wordcloud	80
Gambar 4.35 Hasil Wordcloud Positif.....	80
Gambar 4.36 Hasil Wordcloud Negatif	81
Gambar 4.37 Hasil Wordcloud Netral	81
Gambar 4.38 Script Pie Chart	82
Gambar 4.39 Hasil Pie Chart	82
Gambar L1.1 Profil Obyek Penelitian.....	87
Gambar L2.1 Surat Izin Penelitian.....	88
Gambar L3.1 Balasan Surat Izin Penelitian	89

DAFTAR LAMPIRAN

Lampiran 1. Profil Obyek Penelitian	89
Lampiran 2. Surat Izin Penelitian	90
Lampiran 3. Balasan Surat Izin Penelitian	91



DAFTAR LAMBANG DAN SINGKATAN



TF-IDF	Term Frequency-Inverse Document Frequency
NBC	Naïve Bayes Classifier
API	Application Programming Interface
URL	Uniform Resource Locator
ASCII	American Standard Code for Information Interchange
$\propto$	Proporsionalitas
CSV	Comma Separated Values
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
KBBI	Kamus Besar Bahasa Indonesia
NLP	Natural Language Processing
SVM	Support Vector Machine
LSTM	Long Short-Term Memory
BERT	Bidirectional Encoder Representations from Transformers

DAFTAR ISTILAH

Algoritma	Aturan, tata cara, dan panduan yang dapat digunakan untuk memecahkan sebuah masalah
Klasifikasi	Pengelompokan benda yang sama
Library	Sekumpulan kode yang memiliki fungsi-fungsi tertentu dan dapat dipanggil ke dalam program lain
Script	Program yang ditulis dalam bahasa pemrograman
Term	Kata atau frasa
Frekuensi	Ukuran jumlah terjadinya sebuah peristiwa dalam satuan waktu
Validasi	Pengujian atas kebenaran

INTISARI

Toko Depesaflorist merupakan salah satu toko online yang bergerak di bidang penjualan tanaman hias melalui platform Shopee. Ulasan pelanggan pada toko ini sangat berperan penting dalam membangun reputasi dan kepercayaan calon pembeli. Namun, permasalahan muncul karena banyaknya ulasan yang tidak sesuai dengan jumlah bintang yang diberikan. Hal ini menyulitkan pihak toko dalam memahami opini pelanggan secara menyeluruh dan objektif, apalagi jika jumlah ulasan sangat banyak. Evaluasi manual terhadap ulasan menjadi tidak efisien dan berisiko menimbulkan kesalahan subjektif. Oleh karena itu, diperlukan sistem yang mampu menganalisis dan mengklasifikasikan ulasan secara otomatis berdasarkan sentimen pelanggan.

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi sentimen ulasan produk menggunakan algoritma Naïve Bayes khusus nya varian Multinomial Naïve Bayes. Data ulasan produk pesaing Depesaflorist dikumpulkan melalui proses scraping menggunakan bahasa Python dan API Shopee. Selanjutnya, data menjalani serangkaian tahapan preprocessing seperti cleaning, case folding, normalisasi, tokenizing, stopword removal, dan stemming. Fitur teks kemudian diubah menjadi representasi numerik menggunakan metode pembobotan Term Frequency-Inverse Document Frequency (TF-IDF), sebelum dilakukan klasifikasi sentimen ke dalam tiga kategori yaitu positif, negatif, dan netral.

Hasil klasifikasi menunjukkan bahwa algoritma Multinomial Naïve Bayes mampu mengenali sentimen dalam ulasan dengan akurasi sebesar 52,17% berdasarkan confusion matrix. Model juga menunjukkan performa yang baik dalam mengenali ulasan negatif dengan presisi 73% dan recall 89%. Visualisasi data menggunakan wordcloud dan pie chart turut memperjelas persebaran sentimen pada masing-masing kelas. Sistem yang dikembangkan ini diharapkan dapat membantu pihak toko dalam memahami opini pelanggan, melakukan evaluasi produk, dan meningkatkan kualitas layanan yang ditawarkan.

Kata kunci: Klasifikasi, Ulasan Produk, Naïve Bayes, Marketplace, Analisis Sentiment.

ABSTRACT

Depesaflorist is an online store that sells ornamental plants through the Shopee platform. Customer reviews are essential in shaping the store's reputation and building buyer trust. However, issues arise when the review content does not align with the given star ratings. This inconsistency makes it difficult for the store to objectively understand customer opinions, especially when the volume of reviews is high. Manual evaluation is inefficient and prone to subjective errors. Therefore, an automated system is needed to analyze and classify reviews based on sentiment.

This research aims to develop a sentiment classification system using the Multinomial Naïve Bayes algorithm. Review data were collected from Depesaflorist's competitors via scraping using Python and the Shopee API. The data underwent several preprocessing steps, including cleaning, case folding, normalization, tokenizing, stopword removal, and stemming. Text features were then transformed into numerical form using Term Frequency-Inverse Document Frequency (TF-IDF), followed by sentiment classification into three categories: positive, negative, and neutral.

The classification results show that the Multinomial Naïve Bayes algorithm achieved an accuracy of 52.17% based on the confusion matrix. The model performed well in identifying negative reviews, with 73% precision and 89% recall. Visualization using word clouds and pie charts further clarified sentiment distribution. The proposed system is expected to help the store better understand customer feedback, evaluate products, and improve service quality.

Keyword: Classification, Product Reviews, Naïve Bayes, Marketplace, Sentiment Analysis.