



**ANALISIS *SENTIMENT* PENGGUNA APLIKASI TROVESKIN
BERDASARKAN ULASAN PADA GOOGLE PLAY STORE
MENGUNAKAN METODE NAÏVE BAYES *ALGORITHM***

SKRIPSI

NICODEMUS NAISAU

2010512062

**SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
2024**



**ANALISIS *SENTIMENT* PENGGUNA APLIKASI TROVESKIN
BERDASARKAN ULASAN PADA GOOGLE PLAY STORE
MENGUNAKAN METODE NAÏVE BAYES *ALGORITHM***

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana
Komputer**

NICODEMUS NAISAU

2010512062

**SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
2024**

PERNYATAAN ORISINALITAS

Skripsi Ini adalah hasil karya sendiri dan semua sumber yang dikutip meapun dirujuk telah Saya nyatakan benar. Saya yang menyatakan di bawah ini :

Nama : Nicodemus Naisau

NIM : 2010512062

Studi : S1-Sistem Informasi

Judul : ANALISIS *SENTIMENT* PENGGUNA APLIKASI TROVESKIN BERDASARKAN ULASAN PADA GOOGLE PLAY-STORE MENGGUNAKAN METODE NAÏVE BAYES ALGORITHM

Tanggal : 19 Januari 2024

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan Saya ini, maka Saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 19 Januari 2024

Yang Menyatakan,



(Nicodemus Naisau)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang menandatangani pernyataan ini :

Nama : Nicodemus Naisau
NIM : 2010512062
Fakultas : Ilmu Komputer
Studi : S1- Sistem Informasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**ANALISIS *SENTIMENT* PENGGUNA APLIKASI TROVESKIN
BERDASARKAN ULASAN PADA GOOGLE PLAY STORE MENGGUNAKAN
METODE *NAÏVE BAYES ALGORITHM***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 19 Januari 2024

Yang Menyatakan



(Nicodemus Naisau)

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Dengan ini menyatakan bahwa Skripsi / Tugas Akhir berikut dengan identitas

Nama : Nicodemus Naisau

NIM : 2010512062

Program Studi : S1 - Sistem Informasi

Judul Skripsi / TA : "ANALISIS *SENTIMENT* PENGGUNA APLIKASI TROVESKIN
BERDASARKAN ULASAN PADA GOOGLE PLAY STORE MENGGUNAKAN METODE
NAÏVE BAYES *ALGORITHM*"

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Serjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



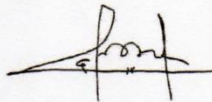
Dr. Widva Cholih, M.I.T

Dosen Pembimbing 1



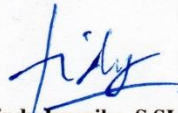
Erly Krisnanik, S.Kom, MM

Dosen Pembimbing 2



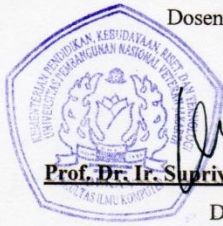
Dr. Ermatita, M.Kom.

Dosen Penguji 1



Nindy Irzavika, S.SI, M.T

Dosen Penguji 2



Prof. Dr. Ir. Suprivanto, ST., M.Sc., IPM

Dekan



Anita Muliawati, S.Kom., MTI.

Ketua Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta

Tanggal Ujian : 12 Januari 2024

ANALISIS *SENTIMENT* PENGGUNA APLIKASI TROVESKIN BERDASARKAN ULASAN PADA GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE BAYES *ALGORITHM*

Nicodemus Naisau

ABSTRAK

TroveSkin merupakan aplikasi *tracking* penggunaan perawatan kecantikan, penelitian ini bertujuan untuk menganalisis *sentiment* pengguna terhadap aplikasi TroveSkin yang berkaitan dengan kualitas performa dan layanan yang ditawarkan. Penelitian ini difokuskan pada evaluasi sentimen pengguna terhadap aplikasi TroveSkin, sebuah platform pelacakan penggunaan perawatan kecantikan. Metodologi penelitian mencakup penggunaan algoritma Naïve Bayes untuk menganalisis sentimen dari 3000 ulasan pengguna di Google Play Store. Data ulasan dilabeli manual berdasarkan skor rating untuk mengklasifikasikan mereka sebagai positif atau negatif. Hasil analisis sentimen digunakan untuk menyusun dashboard visualisasi data, memberikan gambaran yang jelas tentang preferensi dan kepuasan pengguna terhadap performa dan layanan TroveSkin. Hasil penelitian menunjukkan bahwa sentimen pengguna cenderung positif, mencerminkan kepuasan saat ini terhadap layanan aplikasi. Model klasifikasi Naïve Bayes yang diterapkan pada data ulasan mencapai akurasi sebesar 80%, dengan precision 82% dan recall 81%. Penelitian ini memiliki dampak signifikan dalam pemahaman preferensi pengguna terhadap fitur, layanan, dan performa TroveSkin. Rekomendasi berdasarkan analisis ulasan positif termasuk perluasan fitur deteksi kulit, peningkatan kualitas rekomendasi produk, diversifikasi produk, konten edukasi yang lebih kaya, dan fitur jejak perkembangan kulit. Di sisi lain, ulasan negatif menyoroti perbaikan yang diperlukan pada performa aplikasi, responsivitas terhadap umpan balik pengguna, akurasi rekomendasi produk, kejelasan instruksi, dan pertimbangan privasi pengguna.

Kata kunci: *Data Mining, Sentiment Analysis, Naïve Bayes, Dashboard.*

SENTIMENT ANALYSIS OF TROVESKIN APPLICATION USERS BASED ON REVIEWS ON GOOGLE PLAY STORE USING THE NAÏVE BAYES ALGORITHM METHOD

Nicodemus Naisau

ABSTRACT

TroveSkin is a beauty care usage tracking application, and this research aims to analyze user sentiments regarding the TroveSkin app, focusing on the quality of performance and services offered. The research methodology involves the use of the Naïve Bayes algorithm to analyze sentiments from 3000 user reviews on the Google Play Store. Review data is manually labeled based on rating scores to classify them as positive or negative. The results of sentiment analysis are utilized to create a visualized dashboard, providing a clear overview of user preferences and satisfaction with TroveSkin's performance and services. The research findings indicate that user sentiments lean towards the positive, reflecting current satisfaction with the app's services. The Naïve Bayes classification model applied to review data achieves an accuracy of 80%, with precision at 82% and recall at 81%. This research has a significant impact on understanding user preferences regarding TroveSkin's features, services, and performance. Recommendations based on positive reviews include expanding skin detection features, improving the quality of product recommendations, diversifying products, providing richer educational content, and introducing skin progress tracking features. On the other hand, negative reviews highlight the need for improvements in the app's performance, responsiveness to user feedback, accuracy of product recommendations, clarity of instructions, and considerations for user privacy.

Kata kunci: *Data Mining, Sentiment Analysis, Naïve Bayes, Dashboard.*

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, yang telah memberikan Berkah dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi tugas akhir yang berjudul “ANALISIS *SENTIMENT* PENGGUNA APLIKASI TROVESKIN BERDASARKAN ULASAN PADA GOOGLE PLAY STORE MENGGUNAKAN METODE *NAÏVE BAYES ALGORITHM*” Penulis mengucapkan terima kasih yang sebesar-besarnya kepada

1. Tuhan yang Maha Esa atas berkat dan karunia-Nya
2. Orang tua penulis, sebagai pilar utama dalam kehidupan, senantiasa memberikan doa dan dukungan tanpa henti dari proses perkuliahan hingga pengerjaan skripsi ini dan seterusnya.
3. Dosen pembimbing skripsi, Ibu Dr. Widya Cholil, M.I.T dan ibu Erly Krisnanik, S.Kom, MM., atas bimbingan yang telah diberikan dalam setiap tahap pengerjaan dan penyempurnaan skripsi ini.
4. Tidak lupa kepada rekan perkuliahan, Santiana dan Andhika Rizq, yang menjadi teman diskusi berharga sepanjang perkuliahan dan skripsi ini.
5. Mentor Data Management dan rekan rekan yang memberikan *insight* dan juga penyemangat dalam proses pengerjaan skripsi ini.
6. Kepada verifikator *labeling* data, Kevin, Nicholas, dan Della, terima kasih atas kesediaan menjadi bagian penting dalam memastikan validitas data yang digunakan dalam penelitian ini.
7. Kepada rekan-rekan angkatan 2020 Sistem Informasi kelas B, terima kasih atas kerjasama dan dukungan yang telah terjalin dalam perjalanan akademik ini.

Penulis menyadari bahwa tugas akhir ini tidak lepas dari keterbatasan dan kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun untuk pengembangan penelitian di masa yang akan datang.

Jakarta, 22 November 2023

Penulis

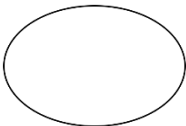


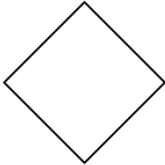


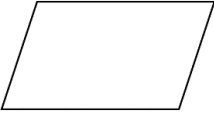
DAFTAR ISI

PERNYATAAN ORISINALITAS.....	i
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS.....	ii
LEMBAR PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR SIMBOL	x
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	6
1.6 Luaran yang Diharapkan	7
1.7 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 Kajian Literatur	9
2.2 Landasan Teori	13
2.2.1 <i>Data Mining</i>	13
2.2.2 <i>Text Mining</i>	13
2.2.3 Analisis Sentimen.....	14
2.2.4 TroveSkin.....	14
2.2.5 Preprocessing Data.....	14
2.2.5.1 Data Cleaning	15
2.2.5.2 Case Folding	15
2.2.5.3 Tokenizing	16
2.2.5.4 Normalization	16
2.2.5.5 Formalization	16
2.2.5.6 Stopwords Removal	16
2.2.5.7 Stemming	17
2.2.6 Labeling Data	17
2.2.7 Ekstraksi Fitur	17

2.2.8 Naïve Bayes Classifier	18
2.2.9 <i>Confusion Matrix</i>	21
2.2.10 Evaluasi Performa	21
2.2.11 <i>K-fold Cross Validation</i>	22
2.2.12 Python	23
2.2.13 Tableau	23
BAB III METODE PENELITIAN	24
3.1 Tempat dan waktu Penelitian	24
3.1.1 Tempat Penelitian.....	24
3.1.2 Waktu Penelitian	24
3.2 Diagram Alur Penelitian.....	25
3.2.1 Identifikasi Masalah	25
3.2.2 Studi Literatur	26
3.2.3 <i>Scraping Data</i>	26
3.2.4 <i>Labeling Data</i>	27
3.2.5 <i>Preprocessing Data</i>	28
3.2.5.1 Data Cleaning	28
3.2.5.2 Case Folding	28
3.2.5.3 Tokenizing	29
3.2.5.4 Normalization	29
3.2.5.5 Formalization	29
3.2.5.6 Stopword Removal	30
3.2.5.7 Stemming	30
3.2.6 <i>Feature Extraction</i>	31
3.2.7 <i>Split Data</i>	31
3.2.8 <i>Training Data</i>	32
3.2.9 <i>Test Data</i>	32
3.2.10 <i>Modeling Naïve Bayes</i>	32
3.2.11 <i>Testing Model Evaluation</i>	33
3.2.12 <i>Performa Evaluation</i>	33
3.2.13 <i>Prediction</i>	33
3.2.14 <i>Dashboard Visualization</i>	34
3.2.15 <i>Analysis & Insight</i>	34
3.3 Alat Bantu Penelitian.....	34
3.4 Jadwal Pelaksanaan.....	35
BAB IV HASIL DAN PEMBAHASAN	37
4.1 Scraping Data	37
4.2 Labeling Data	39
4.3 Preprocessing Data.....	43
4.3.1 Data Cleaning	43
4.3.2 Case Folding.....	44

4.3.3 Tokenizing.....	45
4.3.4 Normalization.....	47
4.3.5 Formalisasi	48
4.3.6 Stopword Removal.....	49
4.3.7 Stemming	49
4.4 Feature Extraction	50
4.5 Split Data	53
4.6 Training Data	53
4.7 Test Data	54
4.8 Modeling Naïve Bayes.....	54
4.9 Testing Model	59
4.10 Performa Evaluation.....	61
4.11 Prediction	63
4.12 Dashboard Visualization	64
4.13 Analysis & Insight.....	66
BAB V PENUTUP	72
5.1 Kesimpulan	72
5.2 Saran	73
DAFTAR PUSTAKA	74
LAMPIRAN.....	79

DAFTAR SIMBOL

No	Simbol Flowchart	Nama	Arti Simbol Flowchart
1		<i>Start</i>	Titik awal atau titik mulai dari alur proses yang akan dijelaskan.
2		<i>Process</i>	Mewakili langkah-langkah atau aktivitas yang dilakukan dalam alur proses.
3		<i>Directional Connector</i>	Menghubungkan dua bagian diagram yang terpisah secara fisik tetapi masih berhubungan dalam aliran proses atau informasi yang sedang dijelaskan.
4		<i>Decision</i>	Mewakili titik pengambilan keputusan dalam aliran proses.
5		<i>Dotted Line</i>	Menunjukkan bahwa aliran atau hubungan antara dua bagian tersebut bersifat opsional, tidak langsung, atau tergantung pada kondisi tertentu
6		<i>Rectangle</i>	Mewakili aktivitas, langkah, atau tindakan dalam aliran proses
7		<i>Parallelogram</i>	Proses atau aktivitas yang berjalan secara parallel atau simultan.

DAFTAR TABEL

Tabel 1.1 Kajian Literatur	9
Tabel 2.2 <i>Confusion Matrix</i>	21
Tabel 3.1 <i>Labeling Data</i> Tahap 1	27
Tabel 3.2 <i>Labeling Data</i> oleh Verifikator.....	27
Tabel 3.3 Perbandingan <i>Data Cleaning</i>	28
Tabel 3.4 Perbandingan <i>Case Folding</i>	28
Tabel 3.5 Perbandingan <i>Tokenizing</i>	29
Tabel 3.6 Perbandingan <i>Normalization</i>	29
Tabel 3.7 Perbandingan Formalisasi	30
Tabel 3.8 Perbandingan <i>Stopwords Removal</i>	30
Tabel 3.9 Perbandingan <i>Stemming</i>	30
Tabel 3.10 Jadwal Penelitian.....	35
Tabel 4.1 Parameter <i>Scraping</i>	37
Tabel 4.2 Distribusi <i>Score</i>	37
Tabel 4.3 <i>Data Type Scraping</i>	38
Tabel 4.4 Hasil Verifikator Sentimen	40
Tabel 4.5 Perbandingan <i>Data Cleaning</i>	43
Tabel 4.6 Perbandingan <i>Case Folding</i>	45
Tabel 4.7 Perbandingan <i>Tokenizing</i>	45
Tabel 4.8 Perbandingan <i>Normalization</i>	47
Tabel 4.9 Perbandingan Formalisasi	48
Tabel 4.10 Perbandingan <i>Stopwords Removal</i>	49
Tabel 4.11 Perbandingan <i>Stemming</i>	50
Tabel 4.12 <i>Term Documents</i>	50
Tabel 4.13 TF-IDF	52
Tabel 4.14 Pembagian <i>Fold Modeling</i>	53
Tabel 4.15 Frekuensi Kata	54
Tabel 4.16 Prediksi Kata Baru	58
Tabel 4.17 Akurasi Modeling Iterasi	60
Tabel 4.18 Performa Model	61
Tabel 4.19 Prediksi Manual Sentimen	63
Tabel 4.20 Prediksi Manual Sentimen ke-2	63

DAFTAR GAMBAR

Gambar 2.1 <i>Cross Validation</i>	22
Gambar 3.1 Diagram Alur Penelitian.....	25
Gambar 4.1 Distribusi <i>Score Users</i>	38
Gambar 4.2 File CSV Hasil <i>Scraping</i>	39
Gambar 4.3 Distribusi Sentimen	40
Gambar 4.4 <i>Anatator Labeling</i>	42
Gambar 4.5 Distribusi Sentimen Verifikator	42
Gambar 4.6 Distribusi Sentimen <i>Modeling</i>	43
Gambar 4.7 <i>Cross Validation Results</i>	60
Gambar 4.8 <i>Confusion Matrix</i>	61
Gambar 4.9 Hasil prediksi Excel	64
Gambar 4.10 <i>Dashboard Home</i>	65
Gambar 4.11 <i>Dashboard Details</i>	65
Gambar 4.12 <i>Pie Chart Distribution Sentiment</i>	66
Gambar 4.13 <i>Trend Line Year - Month Sentimen</i>	66
Gambar 4.14 <i>Word Cloud Sentimen Positive</i>	67
Gambar 4.15 <i>Word Cloud Sentimen Negative</i>	68
Gambar 4.15 <i>Content Ulasan Top Thumbs Up</i>	68