



**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* TERHADAP ANALISIS
SENTIMEN *REVIEW* PENGGUNA APLIKASI *THREADS*, AN
INSTAGRAM APP PADA *GOOGLE PLAY STORE***

SKRIPSI

INDAH FEBRYANA PUTRI

2010512053

PROGRAM STUDI S1 SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

2024



**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* TERHADAP ANALISIS
SENTIMEN REVIEW PENGGUNA APLIKASI *THREADS*, AN
INSTAGRAM APP PADA *GOOGLE PLAY STORE***

SKRIPSI

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

INDAH FEBRYANA PUTRI

2010512053

PROGRAM STUDI S1 SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

2024

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil dari karya sendiri. Semua sumber yang dikutip dan dirujuk telah saya nyatakan dengan benar.

Nama : Indah Febryana Putri

NIM : 2010512053

Program Studi : S1 Sistem Informasi

Judul : IMPLEMENTASI ALGORITMA *NAÏVE BAYES*
TERHADAP ANALISIS SENTIMEN *REVIEW*
PENGGUNA APLIKASITHREADS, AN INSTAGRAM
APP PADA *GOOGLE PLAY STORE*

Tanggal : 19 April 2024

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

Jakarta, 19 April 2024

Yang Menyatakan,

A handwritten signature in black ink is written over a yellow 3000 Rupiah stamp. The stamp features the Garuda Pancasila emblem and the text '3000', 'METERAI KEPOLISIAN', and 'B-12DALX163449514'.

(Indah Febryana Putri)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK
KEPENTINGAN AKADEMIS**

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

Nama : Indah Febryana Putri
NIM : 2010512053
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

**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* TERHADAP ANALISIS
SENTIMEN REVIEW PENGGUNA APLIKASI *THREADS, AN INSTAGRAM APP*
PADA *GOOGLE PLAY STORE***

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 penuli/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di : Jakarta

Pada Tanggal : 19 April 2024

Yang Menyatakan



(Indah Febryana Putri)

LEMBAR PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Indah Febryana Putri

NIM : 2010512053

Program Studi : S1 Sistem Informasi

Judul : IMPLEMENTASI ALGORITMA *NAÏVE BAYES* TERHADAP ANALISIS SENTIMEN *REVIEW* PENGGUNA APLIKASI *THREADS, AN INSTAGRAM APP* PADA *GOOGLE PLAY STORE*

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



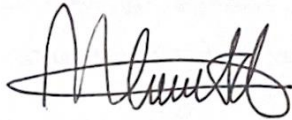
Rio Wirawan, S.Kom., MMSI

Penguji 1



Zatin Niqotaini, S.Tr.Kom., M.Kom

Penguji 2



Nurhafifah Matondang, S.Kom., M.M., M.T.I.

Pembimbing 1



Bambang Tri Wahyono, S.Kom., M.Si

Pembimbing 2



Prof. Dr. P. Supriyanto, ST., M.Sc., IPM

Dekan



Anita Muliawati, S.Kom., MTL

Ketua Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta

Tanggal Sidang : 28 Mei 2024

**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* TERHADAP ANALISIS
SENTIMEN REVIEW PENGGUNA APLIKASI *THREADS*, AN
INSTAGRAM APP PADA *GOOGLE PLAY STORE***

Indah Febryana Putri

ABSTRAK

Di era globalisasi saat ini, media sosial menjadi salah satu produk internet yang paling banyak digunakan oleh masyarakat. Salah satunya adalah *Threads* yang merupakan pendatang baru di industri. *Threads* adalah platform media sosial berbasis teks yang memungkinkan penggunanya untuk berbagi pemikiran, berita, dan konten berbasis teks secara singkat. Dengan tampilan, kegunaan, dan layanannya yang sangat mirip dengan X, banyak pro kontra yang diberikan oleh pengguna terkait aplikasi ini. Oleh karena itu, penulis ingin melakukan penelitian untuk memahami sentimen pengguna *Threads* sehingga dapat memberikan rekomendasi kepada pengembang aplikasi guna meningkatkan performa dan layanan aplikasi. Sebanyak 3322 data ulasan diperoleh dari *Google Play Store*, yang kemudian diberi label positif atau negatif secara manual oleh 3 *annotator*, kemudian masuk ke tahap *preprocessing*, dan pembobotan kata menggunakan metode TF-IDF. Setelah data telah bersih dan memiliki format yang seragam, dilakukan klasifikasi sentimen menggunakan algoritma *Naïve Bayes* dengan pembagian data sebesar 80% untuk *data training* dan 20% untuk *data testing*. Hasil yang diperoleh dari penelitian ini adalah performa dari model klasifikasi *Naïve Bayes* menghasilkan nilai akurasi sebesar 87%. Hasil dari analisis klasifikasi berupa rekomendasi untuk perbaikan teknis dan bug, peningkatan fungsionalitas dan kemudahan pengguna, optimasi konten dan interaksi, serta meningkatkan layanan dan dukungan pengguna.

Kata kunci: Analisis Sentimen, *Threads*, *Naïve Bayes*, *Google Play Store*

**IMPLEMENTATION OF THE NAÏVE BAYES ALGORITHM FOR
SENTIMENT ANALYSIS OF USER REVIEWS OF THREADS, AN
INSTAGRAM APP ON GOOGLE PLAY STORE**

Indah Febryana Putri

ABSTRACT

In the current era of globalization, social media has become one of the most widely used by society. One of them is Threads which is a newcomer to the industry. Threads is a text-based social media platform that allows its users to share thoughts, news and text-based content. With its appearance and services being very similar to X, there are many pros and cons given by users. Therefore, the author wants to conduct research to understand the sentiment of Threads users and provide recommendations to developers to improve the application's performance and services. A total of 3322 reviews was obtained from the Google Play Store, then labeled positive or negative by 3 annotators, entered the preprocessing stage, and feature extraction using TF-IDF. After the data is clean, sentiment classification is carried out using the Naïve Bayes algorithm with a data division of 80% for training data and 20% for testing data. The results are the performance of the Naïve Bayes classification model produces an accuracy value of 87%. The results of the classification analysis are recommendations for technical improvements and bugs, increasing functionality and user convenience, optimizing content and interactions, as well as improving service and user support.

Keywords: *Sentiment Analysis, Threads, Naïve Bayes, Google Play Store*

KATA PENGANTAR

Segala puja dan puji syukur penulis panjatkan atas kehadiran Allah SWT yang telah melimpahkan segala berkat, rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan Tugas Akhir/Skripsi dengan judul “Implementasi Algoritma *Naïve Bayes* Terhadap Analisis Sentimen *Review* Pengguna Aplikasi *Threads*, *An Instagram App* Pada *Google Play Store*” ini dengan tepat waktu. Pada kesempatan ini, penulis ingin mengucapkan banyak terima kasih kepada seluruh pihak yang telah berkontribusi membantu dalam penyusunan Tugas Akhir/Skripsi ini. Terima kasih kepada:

1. Bapak Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM. selaku Dekan Fakultas Ilmu Komputer.
2. Ibu Anita Muliawati, S.Kom., MTI. selaku Ketua Program Studi S1 Sistem Informasi.
3. Ibu Nurhafifah Matondang, S.Kom., M.TI dan Bapak Bambang Triwahyono, S.Kom., M.Si selaku dosen pembimbing 1 dan dosen pembimbing 2 penulis.
4. Orang tua dan keluarga penulis yang senantiasa mendoakan dan mendukung baik secara materi dan non-materi.
5. Teman-teman seperjuangan penulis dan seluruh pihak yang tidak bisa penulis sebutkan satu persatu yang senantiasa memberikan dukungan kepada penulis dengan sepenuh hati.

Akhir kata, penulis berharap semoga Skripsi ini dapat bermanfaat dan menambah pengetahuan atau pun wawasan kepada para pembaca. Penulis menyadari bahwa Skripsi ini jauh dari kata sempurna, dari segi pengetahuan, penulisan, maupun penyampaiannya. Penulis akan sangat mengapresiasi dan menerima kritik dan saran yang dapat membangun agar Skripsi ini dapat tersusun lebih baik.

Jakarta, 19 April 2024

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 GAMBAR	x
DAFTAR TABEL	xi
DAFTAR RUMUS	xii
DAFTAR SIMBOL	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Ruang Lingkup.....	5
1.5 Manfaat Penelitian	6
1.6 Luaran Yang Diharapkan	6
1.7 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	8
2.1 <i>Threads</i>	8
2.2 <i>Algoritma Naïve Bayes</i>	9
2.2.1 <i>Naïve Bayes Classifier</i>	9
2.2.2 <i>Multinomial Naïve Bayes</i>	10
2.3 Analisis Sentimen	11
2.4 <i>Data Mining</i>	12
2.5 <i>Text Mining</i>	13
2.6 <i>Annotator</i>	14
2.7 <i>Data Pre-processing</i>	14
2.7.1 <i>Data Cleansing</i>	14
2.7.2 <i>Normalization</i>	14

2.7.3	<i>Stopword Removal</i>	15
2.7.4	<i>Stemming</i>	15
2.7.5	<i>Tokenizing</i>	15
2.8	Ekstraksi Fitur	16
2.9	<i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	16
2.10	Confusion Matrix	18
2.11	Evaluasi Performa.....	19
2.11.1	<i>Accuracy</i>	19
2.11.2	<i>Precision</i>	19
2.11.3	<i>Recall</i>	19
2.11.4	<i>Specificity</i>	20
2.11.5	<i>F1 Score</i>	20
2.12	Python	20
2.13	Penelitian Terdahulu	21
2.14	Hipotesis	25
BAB III METODE PENELITIAN		26
3.1	Alur Penelitian	26
3.1.1	Identifikasi Masalah.....	27
3.1.2	Studi Literatur	27
3.1.3	Pengumpulan Data	28
3.1.4	Pelabelan Data.....	28
3.1.5	<i>Data Pre-processing</i>	29
3.1.6	Pembobotan Kata	34
3.1.7	Pembagian Data	35
3.1.8	Pengujian Model <i>Naïve Bayes</i>	36
3.1.9	Evaluasi Performa.....	37
3.1.10	Analisis Hasil Klasifikasi.....	37
3.1.11	Visualisasi	37
3.2	Perangkat Pendukung Penelitian.....	38
3.3	Jadwal Penelitian	39
BAB IV HASIL DAN PEMBAHASAN		41
4.1	Pengumpulan Data	41
4.2	Pelabelan Data	41
4.3	<i>Data Pre-processing</i>	43
4.3.1	<i>Data Cleansing</i>	43

4.3.2	<i>Normalization</i>	44
4.3.3	<i>Stopword Removal</i>	46
4.3.4	<i>Stemming</i>	47
4.3.4	<i>Tokenizing</i>	48
4.4	Pembobotan Kata	48
4.5	Pembagian Data	52
4.6	Pengujian Model <i>Naïve Bayes</i>	52
4.7	Evaluasi Performa.....	58
4.8	Analisis Hasil Klasifikasi.....	60
4.9	Visualisasi.....	68
BAB V PENUTUP		73
5.1	Kesimpulan	73
5.2	Saran	73
DAFTAR PUSTAKA		75
DAFTAR RIWAYAT HIDUP		80
LAMPIRAN		81

DAFTAR GAMBAR

Gambar 1.1 Jumlah Pengguna Internet di Indonesia (Sumber: <i>We Are Social</i> , 2024)	1
Gambar 3.1 Alur Kerja Penelitian.....	26
Gambar 3.2 Dataset.....	28
Gambar 4.1 Hasil <i>Scraping Data</i>	41
Gambar 4.2 Diagram Pie Distribusi Sentimen.....	61
Gambar 4.3 Grafik Distribusi Panjang Ulasan Sebelum <i>Preprocessing</i>	61
Gambar 4.4 Grafik Distribusi Panjang Ulasan Setelah <i>Preprocessing</i>	62
Gambar 4.5 Distribusi Jumlah <i>Review</i> per bulan.....	62
Gambar 4.6 30 Kata Terbanyak secara Keseluruhan.....	63
Gambar 4.7 <i>Word Cloud</i> Keseluruhan Data Ulasan	63
Gambar 4.8 30 Kata Terbanyak dengan Label Positif.....	64
Gambar 4.9 <i>Word Cloud</i> Data Ulasan Positif.....	64
Gambar 4.10 30 Kata Terbanyak dengan Label Negatif.....	66
Gambar 4.11 <i>Word Cloud</i> Data Ulasan Negatif.....	66
Gambar 4.12 Halaman <i>Dashboard Web</i> (1).....	69
Gambar 4.13 Halaman <i>Dashboard Web</i> (2).....	69
Gambar 4.14 Halaman <i>Word Cloud Web</i>	70
Gambar 4.15 Fitur <i>Dropdown</i> halaman <i>Word Cloud</i>	70
Gambar 4.16 Halaman <i>Trends Web</i> (1)	71
Gambar 4.17 Halaman <i>Trends Web</i> (2)	71
Gambar 4.18 Fitur <i>Dropdown</i> Halaman <i>Trends</i>	72

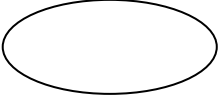
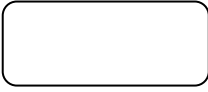




DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	18
Tabel 2.2 Penelitian Terdahulu	21
Tabel 2.3 Hipotesis.....	25
Tabel 3.1 Proses Pelabelan Data	29
Tabel 3.2 Proses <i>Data Cleaning</i>	30
Tabel 3.3 Proses <i>Case Folding</i>	31
Tabel 3.4 Proses <i>Normalization</i>	31
Tabel 3.5 Proses <i>Stopword Removal</i>	32
Tabel 3.6 Proses <i>Stemming</i>	33
Tabel 3.7 Proses <i>Tokenizing</i>	33
Tabel 3.8 Contoh Data	34
Tabel 3.9 Proses Perhitungan TF-IDF	34
Tabel 3.10 Detail Pembagian Data.....	35
Tabel 3.11 Jadwal Penelitian.....	39
Tabel 4.1 Contoh Pelabelan Data.....	42
Tabel 4.2 Hasil Pelabelan Data	43
Tabel 4.3 Contoh Data sebelum <i>Preprocessing</i>	43
Tabel 4.4 Proses <i>Data Cleansing</i>	44
Tabel 4.5 Proses <i>Normalization</i>	45
Tabel 4.6 Proses <i>Stopword Removal</i>	46
Tabel 4.7 Proses <i>Stemming</i>	47
Tabel 4.8 Proses <i>Tokenizing</i>	48
Tabel 4.9 Dokumen Ulasan.....	49
Tabel 4.10 Proses Pembobotan Kata.....	50
Tabel 4.11 Rincian Pembagian Data.....	52
Tabel 4.12 Sampel <i>Data Testing</i> sebelum <i>Preprocessing</i>	55
Tabel 4.13 Sampel <i>Data Testing</i> setelah <i>Preprocessing</i>	55
Tabel 4.14 Perhitungan TF-IDF untuk <i>Data Testing</i>	55
Tabel 4. 15 Hasil <i>Confusion Matrix</i> Model <i>Naïve Bayes</i>	59

DAFTAR RUMUS

Rumus <i>Multinomial Naïve Bayes</i> (1)	10
Rumus <i>Prior Probability</i> (2)	10
Rumus <i>Multinomial Naïve Bayes</i> dengan TF-IDF (3)	11
Rumus TF (4)	17
Rumus IDF (5)	17
Rumus TF-IDF (6)	17
Rumus <i>Accuracy</i> (7)	19
Rumus <i>Precision</i> (8)	19
Rumus <i>Recall</i> (9)	20
Rumus <i>Specificity</i> (10)	20
Rumus <i>F1 Score</i> (11)	20

DAFTAR SIMBOL

No.	Simbol <i>Flowchart</i>	Nama	Deskripsi
1.		<i>Start</i>	Titik mulai atau titik awal dari alur proses yang akan dijelaskan.
2.		<i>Process</i>	Merepresentasikan tahapan-tahapan atau aktivitas yang dilakukan dalam alur proses.
3.		<i>Rectangle</i>	Mewakili sebuah aktivitas, proses, atau langkah dalam sebuah aliran pada proses.
4.		<i>Parallelogram</i>	Digunakan untuk menandakan bahwa proses atau aktivitas berjalan secara parallel atau bersamaan.
5.		<i>Directional Connector</i>	Penghubung antara dua segmen diagram yang terpisah secara fisik, namun tetap terhubung dalam alur proses atau informasi yang sedang dijelaskan.
6.		<i>Dotted Line</i>	Menunjukkan bahwa hubungan antara dua elemen tersebut bersifat pilihan (opsional), tidak langsung, atau bergantung pada situasi khusus.