



**VISUALISASI HASIL ANALISIS SENTIMEN PENGGUNA APLIKASI PAHAMIFY
BERDASARKAN ULASAN PADA GOOGLE PLAY STORE MENGGUNAKAN
METODE ALGORITMA NAÏVE BAYES BERBASIS WEB**

SKRIPSI

**RAFFAEL
2010512058**

**PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
2024**



**VISUALISASI HASIL ANALISIS SENTIMEN PENGGUNA APLIKASI PAHAMIFY
BERDASARKAN ULASAN PADA GOOGLE PLAY STORE MENGGUNAKAN
METODE ALGORITMA NAÏVE BAYES BERBASIS WEB**

SKRIPSI

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

RAFFAEL

2010512058

**PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
2024**

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Raffael

NIM : 2010512058

Tanggal : 28 April 2024

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

Jakarta, 28 April 2024

Yang Menyatakan,



(Raffael)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI
UNTUK KEPENTINGAN AKADEMIS**

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN
AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Raffael

NIM : 2010512058

Fakultas : Ilmu Komputer

Program 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:

Visualisasi Hasil Analisis Sentimen Pengguna Aplikasi Pahamify Berdasarkan Ulasan pada Google Play Store Menggunakan Metode Algoritma *Naïve Bayes* Berbasis Web

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih data/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Dibuat di : Jakarta

Pada tanggal : 28 April 2024

Yang Menyatakan,



Raffael

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Raffael
NIM : 2010512058
Program Studi : S1 – Sistem Informasi

Judul Tugas Akhir : Visualisasi Hasil Analisis Sentimen Pengguna Aplikasi Pahamify Berdasarkan Ulasan Pada Google Play Store Menggunakan Metode Algoritma Na'Ve Bayes Berbasis Web.

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



Rio Wirawan, S.Kom., MMSI.

Penguji 1



Zatin Nigotaini, S.Tr.Kom., M.Kom.

Penguji 2



Nur Hafifah Matondang, S.Kom., M.M., M.T.I.

Pembimbing 1



Ery Krisnanik, S.Kom., MM

Pembimbing 2



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

Dekan



Anita Muliawati, S.Kom., MTI.

Ketua Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta

Tanggal Ujian : 24 April 2024

VISUALISASI HASIL ANALISIS SENTIMEN PENGGUNA APLIKASI PAHAMIFY BERDASARKAN ULASAN PADA GOOGLE PLAY STORE MENGGUNAKAN METODE ALGORITMA NAÏVE BAYES BERBASIS WEB

Raffael

ABSTRAK

Pembelajaran daring telah menjadi salah satu kemajuan teknologi dalam pendidikan, terutama dalam menghadapi tantangan pandemi COVID-19. Salah satu aplikasi pembelajaran daring yang populer adalah Pahamify. Pahamify menawarkan berbagai fitur untuk membantu siswa mempersiapkan diri untuk ujian dan meningkatkan pemahaman materi. Namun, seperti halnya dengan banyak aplikasi, Pahamify juga memiliki kelemahan, termasuk rendahnya minat belajar siswa dan kesulitan dalam memahami materi secara daring. Evaluasi kualitasnya dapat dilakukan melalui analisis sentimen yaitu sebuah teknik untuk mengidentifikasi dan memahami pendapat atau sentimen dari pengguna berdasarkan ulasan yang diberikan. Penelitian ini menggunakan metode Naïve Bayes Classifier untuk menganalisis sentimen pengguna terhadap aplikasi Pahamify berdasarkan ulasan di Google Play Store. Data ulasan yang dikumpulkan dimulai dari Agustus 2020 hingga Mei 2022 kemudian dilabeli secara manual oleh beberapa annotator. Data akan melalui berbagai tahapan sebelum dilakukannya klasifikasi seperti tahap *preprocessing*, pembobotan kata dengan metode *Term Frequency – Inverse Document Frequency* (TF-IDF), dan pembagian data. Hasil penelitian menunjukkan bahwa model klasifikasi Naïve Bayes menghasilkan akurasi yang baik, dengan nilai akurasi mencapai 86%, recall sebesar 95%, dan precision sebesar 84% dengan menggunakan rasio data latih dan uji 90:10. Selain hasil klasifikasi, penelitian juga membuat visualisasi data berupa *Word Cloud* untuk mengidentifikasi kata-kata kunci yang sering muncul dalam ulasan positif dan negatif. Luaran lainnya yang diperoleh sebuah sistem sederhana untuk memprediksi label berdasarkan data ulasan yang diberikan.

Kata Kunci : Analisis Sentimen, *Google Play Store*, *Naïve Bayes*, Pahamify

**VISUALIZATION SENTIMENT ANALYSIS RESULTS OF PAHAMIFY APPLICATION
USERS BASED ON REVIEWS GOOGLE PLAY STORE USING WEB-BASED NAÏVE
BAYES ALGORITHM METHOD**

Raffael

ABSTRACT

Online learning has become one of the technological advances in education, especially in facing the challenges of the COVID-19 pandemic. One of the popular online learning apps is Pahamify. Pahamify offers various features to help students prepare for exams and improve understanding of the material. However, as is the case with many apps, Pahamify also has weaknesses, including low student interest in learning and difficulty in understanding the material online. Evaluation of its quality can be done through sentiment analysis, which is a technique to identify and understand the opinions or sentiments of users based on the reviews given. This research uses the Naïve Bayes Classifier method to analyze user sentiment towards the Pahamify application based on reviews on the Google Play Store. The review data collected from August 2020 to May 2022 was then manually labeled by several annotators. The data will go through various stages before classification such as preprocessing, word weighting with the Term Frequency - Inverse Document Frequency (TF-IDF) method, and data division. The results show that the Naïve Bayes classification model produces good accuracy, with accuracy values reaching 86%, recall of 95%, and precision of 84% using a training and test data ratio of 90:10. In addition to the classification results, the research also created a data visualization in the form of Word Cloud to identify key words that often appear in positive and negative reviews. Another output obtained is a simple system for predicting labels based on the given review data.

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

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, yang telah memberikan kasih karunia dan berkat-Nya sehingga penulis dapat berkesempatan untuk menyelesaikan Tugas Akhir yang berjudul “VISUALISASI HASIL ANALISIS SENTIMEN PENGGUNA APLIKASI PAHAMIFY BERDASARKAN ULASAN PADA GOOGLE PLAY STORE MENGGUNAKAN METODE ALGORITMA NAÏVE BAYES BERBASIS WEB” dengan baik. Penelitian ini ditempuh karena merupakan persyaratan untuk pengambilan Tugas Akhir/Skripsi. Skripsi ini dapat selesai tidak lepas dari bantuan banyak pihak yang telah turut serta dalam memberikan bantuan dan bimbingannya kepada saya. Oleh karena itu, penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM, selaku dekan Fakultas Ilmu Komputer.
2. Ibu Anita Muliawati, S.Kom., MTI selaku Ketua Program Studi Sarjana Sistem Informasi.
3. Ibu Nur Hafifah Matondang, S.Kom., MM., M.T.I selaku Dosen Pembimbing 1.
4. Ibu Erly Krisnanik, S.Kom.,MM selaku Dosen Pembimbing 2.
5. Seluruh keluarga terutama kedua orang tua yang selalu memberikan dukungan baik secara moral maupun materi.
6. Seluruh kawan dan teman-teman terutama Bagas, Keluarga Besar Alfito, dan Petrus yang telah selalu memberikan dukungan dan semangat.
7. Seluruh pihak yang tidak bisa penulis sebutkan satu-persatu.

Peneliti menyadari memiliki banyak kekurangan yang ada pada Proposal Tugas Akhir ini, baik dari materi hingga teknik penulisan . Oleh karena itu, peneliti menerima kritik dan saran membangun agar dapat lebih baik untuk selanjutnya.

Jakarta, 14 Maret 2024

Penulis

DAFTAR ISI

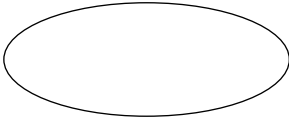

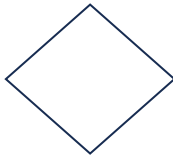

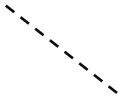
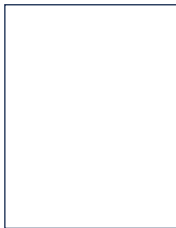
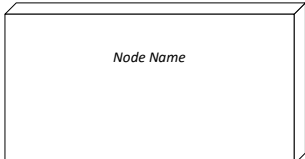
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
ABSTRAK.....	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR SIMBOL	xiii
DAFTAR RUMUS	xv
DAFTAR TABEL.....	xvii
DAFTAR GAMBAR.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Ruang Lingkup Penelitian.....	5
1.6 Luaran yang diharapkan	5
1.7 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Analisis Sentimen.....	7
2.2 <i>Web Scraping</i>	7
2.3 <i>Google Play Store</i>	8
2.4 Pahamify.....	8
2.5 Sistem	8
2.6 <i>Use Case Diagram</i>	8
2.7 <i>Deployment Diagram</i>	9
2.8 <i>Python</i>	9
2.9 <i>Data Mining</i>	9
2.9.1 <i>Text Mining</i>	10


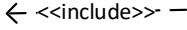
2.10	<i>Preprocessing</i>	10
2.10.1	<i>Case Folding</i>	10
2.10.2	<i>Data Cleaning</i>	10
2.10.3	<i>Normalization</i>	11
2.10.4	<i>Stemming</i>	11
2.10.5	<i>Stopword Removal</i>	11
2.10.6	<i>Tokenization</i>	11
2.11	Pembobotan Kata.....	11
2.12	Klasifikasi.....	13
2.13	<i>Naive Bayes Classifier</i>	13
2.14	Evaluasi Model Klasifikasi.....	16
2.15	Annotator.....	16
2.16	<i>Fleiss Kappa</i>	17
2.17	<i>Black-Box Testing</i>	18
2.18	<i>Flask</i>	18
2.19	Pendahuluan Terdahulu.....	19
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	Tahapan Penelitian.....	25
3.2.1	Identifikasi Masalah.....	25
3.2.2	Studi Literatur.....	26
3.2.3	Pengumpulan Data.....	26
3.2.4	Pelabelan Data.....	26
3.2.5	<i>Preprocessing</i>	28
3.2.6	Pembobotan TF-IDF.....	32
3.2.7	Pembagian Data.....	32
3.2.8	Klasifikasi Menggunakan <i>Multinomial Naive Bayes</i>	33
3.2.9	Analisis dan Evaluasi Algoritma <i>Multinomial Naive Bayes</i>	33
3.2.10	Analisis Hasil Klasifikasi dan Visualisasi.....	34
3.2.11	Perancangan Sistem.....	35
3.2.12	Implementasi dan Pengujian Sistem.....	35

3.3	Alat Bantu Penelitian	35
3.4	Jadwal Penelitian.....	36
BAB IV HASIL DAN PEMBAHASAN		38
4.1	Pengumpulan Data	38
4.2	Pelabelan Data.....	39
4.3	<i>Preprocessing</i>	44
4.3.1	<i>Case Folding</i>	44
4.3.2	<i>Data Cleaning</i>	44
4.3.3	<i>Normalization</i>	45
4.3.4	<i>Stemming</i>	45
4.3.5	<i>Stopword Removal</i>	46
4.3.6	<i>Tokenization</i>	46
4.4	Pembobotan TF-IDF	47
4.5	Pembagian Data.....	49
4.6	Klasifikasi Naïve Bayes	50
4.7	Analisis dan Evaluasi Algoritma Naïve Bayes	56
4.7.1	Evaluasi Model <i>Naïve Bayes</i> dengan Rasio Data Pertama	57
4.7.2	Evaluasi Model <i>Naïve Bayes</i> dengan Rasio Data Kedua.....	58
4.7.3	Evaluasi Model <i>Naïve Bayes</i> dengan Rasio Data Ketiga.....	60
4.7.4	Evaluasi Perbandingan Hasil Model <i>Multinomial Naïve Bayes</i>	61
4.8	Analisis Hasil Klasifikasi dan Visualisasi.....	62
4.8.1	Visualisasi <i>Word Cloud</i> Sentimen	62
4.8.2	Visualisasi <i>Word Cloud</i> Sentimen Positif.....	63
4.8.3	Visualisasi <i>Word Cloud</i> Sentimen Negatif	63
4.9	Perancangan Sistem.....	64
4.9.1	<i>Use Case Diagram</i>	65
4.9.2	<i>Deployment Diagram</i>	65
4.10	Implementasi dan Pengujian Sistem	66
4.10.1	Tampilan Halaman <i>Landing Page</i>	66
4.10.2	Tampilan Halaman Prediksi Sentimen Analisis	67
4.10.3	Tampilan Halaman Visualisasi Sentimen Analisis	68
4.10.4	Tampilan Halaman <i>About</i>	68
4.10.5	Pengujian dengan <i>Black-Box Testing</i>	69

BAB V PENUTUP	71
5.1 Kesimpulan.....	71
5.2 Saran.....	71
DAFTAR PUSTAKA	72
DAFTAR RIWAYAT HIDUP	75
LAMPIRAN.....	76

DAFTAR SIMBOL

No	Simbol	Nama	Keterangan
1.		Terminator	Simbol yang menyatakan sebuah awal maupun akhir dari sebuah alur program.
		<i>Process</i>	Simbol yang menyatakan suatu proses atau langkah yang berjalan sepanjang alur program.
3.		<i>Decision</i>	Simbol yang mewakili pengambilan keputusan untuk langkah selanjutnya pada alur program.
4.		<i>Flow Line</i>	Simbol yang menggabungkan antara satu simbol dengan simbol lainnya yang mempunyai keterhubungan pada alur program.
5.		<i>Dotted Line</i>	Simbol yang menyatakan alur program dengan keterhubungan langkah yang bisa bersifat opsional atau bergantung pada kondisi tertentu.
6.		<i>System</i>	Simbol yang menyatakan sebuah sistem dan interaksi yang dapat dilakukan pada sistem serta membatasi interaksi dari luar sistem.
7.		<i>Node</i>	Simbol ini biasanya mewakili suatu <i>hardware</i> (perangkat keras) atau <i>software</i> (perangkat lunak).

No	Simbol	Nama	Keterangan
8.		<i>Actor</i>	Simbol yang mewakili peran user atau pengguna ketika berinteraksi dengan sebuah sistem.
9.		<i>Include</i>	Simbol yang menunjukkan suatu <i>use case</i> seluruhnya adalah fungsionalitas dari <i>use case</i> lainnya.

DAFTAR RUMUS

No	Rumus	Keterangan
(1)	$TF = Wd_v t_j$	Rumus dalam menghitung TF
(2)	$IDF = \ln\left(\frac{N}{df_t}\right) + 1$	Rumus dalam menghitung IDF
3.	$TF\ IDF = TF \times IDF$	Rumus dalam menghitung TF-IDF
4.	$P(H X) = \frac{P(X H) P(H)}{P(X)}$	Rumus Teorema Naïve Bayes
5.	$P_{(c)} = \frac{N_c}{N_{doc}}$	Rumus menghitung nilai probabilitas pada tiap kelas
6.	$P(Wi c) = \frac{count(w_{ij,c})}{ c + V }$	Rumus menghitung nilai probabilitas suatu kata pada tiap kelas
7.	$P(Wi c) = \frac{count(w_{ij,c}) + 1}{ c + V }$	Rumus menghitung nilai probabilitas suatu kata menggunakan <i>Laplace Smoothing</i>
8.	$C_{MAP} = \underset{c \in V}{\arg \max} P(c) \prod_{t=1}^{ V } P(Wi c)$	Rumus menghitung <i>Multinomial Naïve Bayes</i>
(9)	$Kappa\ Value = \frac{p_0 - p_e}{1 - p_e}$	Rumus Persamaan <i>Kappa Value</i>
(10)	$P_0 = \frac{1}{n} \sum_{i=1}^n q_i$	Rumus perhitungan persentase jumlah pengukuran antar <i>rater</i>

No	Rumus	Keterangan
(11)	$P_e = \sum_{j=1}^k q_j^2$	Rumus perhitungan persentase jumlah perubahan antar <i>rater</i>
(12)	$q_i = \frac{1}{m(m-1)} \left(\sum_{j=1}^k m_{ij}^2 - m_{ij} \right)$	Rumus mencari nilai q_i untuk mendapatkan P_0
(13)	$q_j = \frac{m \cdot j}{nm} = \frac{1}{nm} \sum_{i=1}^n m_{ij}$	Rumus mencari nilai q_j untuk mendapatkan P_e
(14)	$\frac{TP + TN}{TP + TN + FP + FN}$	Rumus menghitung <i>accuracy</i> klasifikasi <i>Naïve Bayes</i>
(15)	$\frac{TP}{TP + FP}$	Rumus menghitung <i>precision</i> klasifikasi <i>Naïve Bayes</i>
(16)	$\frac{TP}{TP + FN}$	Rumus menghitung <i>recall</i> klasifikasi <i>Naïve Bayes</i>

DAFTAR TABEL

Tabel 2.1 Skala Nilai <i>Kappa</i>	18
Tabel 2.2 Penelitian Terdahulu.....	19
Tabel 3.1 Contoh Pengumpulan Data	26
Tabel 3.2 Contoh Pelabelan Data.....	27
Tabel 3.3 Contoh proses <i>case folding</i>	28
Tabel 3.4 Contoh proses <i>data cleaning</i>	29
Tabel 3.5 Contoh proses <i>normalization</i>	29
Tabel 3.6 Contoh proses <i>stemming</i>	30
Tabel 3.7 Contoh proses <i>stopword removal</i>	31
Tabel 3.8 Contoh proses <i>tokenization</i>	31
Tabel 3.9 Contoh Persentase Pembagian Data.....	32
Tabel 3.10 Rumus <i>Confusion Matrix</i>	34
Tabel 3.11 Jadwal Penelitian.....	36
Tabel 4.1 Hasil Pelabelan pada Ulasan	39
Tabel 4.2 Hasil Pelabelan oleh <i>Annotator</i>	41
Tabel 4.3 Hasil Perhitungan Nilai q_i Setiap Dokumen	42
Tabel 4. 4 Tahapan <i>Preprocessing Case Folding</i>	44
Tabel 4. 5 Tahapan <i>Preprocessing Data Cleaning</i>	44
Tabel 4.6 Tahapan <i>Preprocessing Normalization</i>	45
Tabel 4.7 Tahapan <i>Preprocessing Stemming</i>	45
Tabel 4.8 Tahapan <i>Preprocessing Stopword Removal</i>	46
Tabel 4.9 Tahapan <i>Preprocessing Tokenization</i>	46
Tabel 4.10 Data Sampel Perhitungan TF-IDF	47
Tabel 4.11 Hasil Perhitungan Pembobotan TF-IDF.....	48
Tabel 4.12 Pembagian Data dengan Rasio 90 : 10.....	49
Tabel 4.13 Tabel Pembagian Data dengan Rasio 80 : 20.....	49
Tabel 4.14 Tabel Pembagian Data dengan Rasio 70 : 30.....	49
Tabel 4.15 Sampel Data Latih <i>Multinomial Naïve Bayes</i>	50
Tabel 4.16 Bobot Kata dari Sampel Data Latih	50
Tabel 4.17 Nilai Seluruh Probabilitas Data Latih <i>Multinomial Naïve Bayes</i>	53

Tabel 4.18 Sampel Data Uji <i>Multinomial Naïve Bayes</i>	54
Tabel 4.19 Perhitungan TF-IDF Sampel Data Uji <i>Multinomial Naïve Bayes</i>	54
Tabel 4.20 Perbandingan <i>Accuracy</i> Model <i>Naïve Bayes</i>	61
Tabel 4.21 <i>Test case</i> Web Sistem Analisis	69

DAFTAR GAMBAR

Gambar 1.1 Google Play Store Categories Q3 2022 dari artikel Statista	2
Gambar 3.1 Tahapan Penelitian	25
Gambar 4.1 Hasil <i>Scraping</i> Ulasan Pahamify dari <i>Google Play Store</i>	38
Gambar 4.2 Hasil <i>Confusion Matrix</i> dengan Rasio data 70 : 30	57
Gambar 4.3 Hasil <i>Confusion Matrix</i> dengan Rasio data 80 : 20	59
Gambar 4.4 Hasil <i>Confusion Matrix</i> dengan Rasio data 90 : 10	60
Gambar 4.5 <i>Word Cloud</i> Label Positif dan Negatif	62
Gambar 4.6 <i>Word Cloud</i> Label Positif.....	63
Gambar 4.7 <i>Word Cloud</i> Label Negatif	64
Gambar 4. 8 Rancangan <i>Use Case</i> Diagram Sistem Sederhana	65
Gambar 4. 9 Rancangan <i>Deployment</i> Diagram Sistem Sederhana	65
Gambar 4.10 Tampilan Laman <i>Landing Page</i>	66
Gambar 4.11 Tampilan Laman Prediksi Sentimen Analisis.....	67
Gambar 4.12 Tampilan Laman Visualisasi Sentimen Analisis	68
Gambar 4.13 Tampilan Laman <i>About</i>	68