

**ANALISIS SENTIMEN ULASAN PADA APLIKASI HAJI PINTAR  
MENGUNAKAN ALGORITMA NAÏVE BAYES**



Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Muhammad Rezalutfi**

**0000028098**

**UMMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2023**

**ANALISIS SENTIMEN ULASAN PADA APLIKASI HAJI PINTAR  
MENGUNAKAN ALGORITMA NAÏVE BAYES**



Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Muhammad Rezalutfi**

**00000028098**

**UMMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2023**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

### HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Muhammad Rezalutfi  
Nomor Induk Mahasiswa : 00000028098  
Program Studi : Informatika

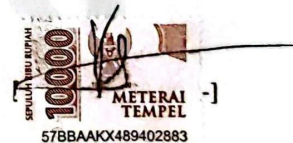
Skripsi dengan judul:

**Analisis Sentimen Ulasan Pada Aplikasi Haji Pintar Menggunakan Algoritma Naïve Bayes**

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 16 Juni 2023



(Muhammad Rezalutfi)

## HALAMAN PENGESAHAN

### HALAMAN PENGESAHAN

Skripsi dengan judul

#### **ANALISIS SENTIMEN ULASAN PADA APLIKASI HAJI PINTAR MENGUNAKAN ALGORITMA NAÏVE BAYES**

oleh

Nama : Muhammad Rezalutfi  
NIM : 00000028098  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika


Telah diujikan pada hari Selasa, 20 Juni 2023

Pukul 10.00 s/s 12.00 dan dinyatakan

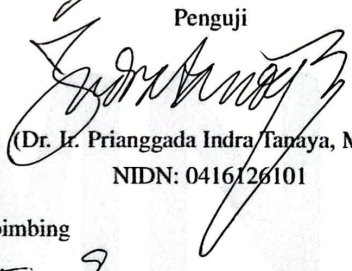
**LULUS**

Dengan susunan penguji sebagai berikut


Ketua Sidang

  
(Angga Aditya Permana, S.Kom.,  
M.Kom.)  
NIDN: 0407128901

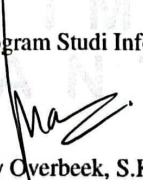
Penguji

  
(Dr. Ir. Prianggada Indra Tanaya, MME)  
NIDN: 0416126101

Pembimbing

  
(Adhi Kusnadi, S.T, M.Si.)  
NIDN: 303037304

Ketua Program Studi Informatika,

  
(Marlinda Vasty Overbeek, S.Kom., M.Kom.)  
NIDN: 0818038501

iii

Analisis Sentimen Ulasan..., Muhammad Rezalutfi, Universitas Multimedia Nusantara

iii

Analisis Sentimen Ulasan..., Muhammad Rezalutfi, Universitas Multimedia Nusantara

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

### HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Rezalutfi  
NIM : 0000028098  
Program Studi : Informatika  
Fakultas : Teknik dan Informatika  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

#### ANALISIS SENTIMEN ULASAN PADA APLIKASI HAJI PINTAR MENGUNAKAN ALGORITMA NAÏVE BAYES

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

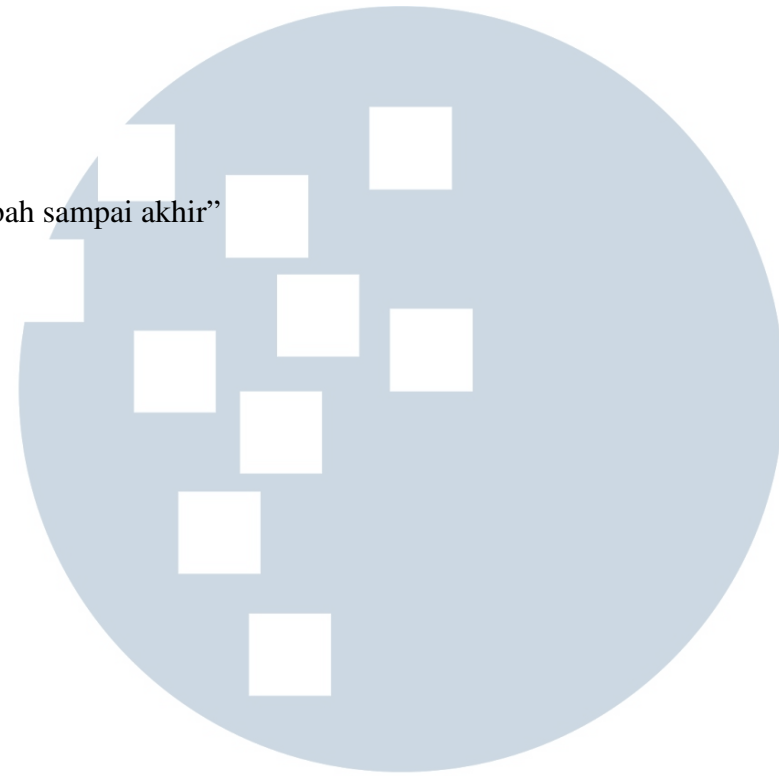
Tangerang, 16 Juni 2023  
Yang menyatakan



Muhammad Rezalutfi

**Halaman Persembahan / Motto**

”Tabah sampai akhir”



**UMMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis Sentimen Ulasan Pada Aplikasi Haji Pintar Menggunakan Algoritma Naïve Bayes dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Adhi Kusnadi, S.T, M.Si., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua, Keluarga, dan teman-teman saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.
6. Kepada diri saya Muhammad Rezalutfi yang selalu tabah sampai akhir.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi bagi para pembaca.

Tangerang, 16 Juni 2023



Muhammad Rezalutfi

# ANALISIS SENTIMEN ULASAN PADA APLIKASI HAJI PINTAR MENGUNAKAN ALGORITMA NAÏVE BAYES

Muhammad Rezalutfi

## ABSTRAK

Aplikasi Haji Pintar merupakan aplikasi yang digunakan oleh para jamaah haji di Indonesia untuk mendapatkan informasi terkait pelaksanaan haji dan membantu jamaah saat melakukan ibadah Haji. Beragam macam ulasan pengguna aplikasi Haji Pintar pada Google Play Store mengenai keluhan dan saran terhadap aplikasi, oleh karena itu penelitian dilakukan untuk menganalisis sentimen positif, netral, dan negatif dari ulasan pengguna aplikasi Haji Pintar pada Google Play Store. Penelitian ini menggunakan metode *Text Preprocessing* yang terdiri dari *Data Cleaning*, *Case Folding*, *Tokenization*, *Stopword Removal*, *Stemming*, yang dilanjutkan dengan pembobotan *Term Frequency - Inverse Document Frequency*, kemudian model klasifikasi algoritma Naïve Bayes dengan *data split* dan *K-Fold Cross Validation*. Hasil Pengujian terbaik didapatkan oleh model klasifikasi algoritma Naïve Bayes menggunakan *Data Split* dengan 80:20 yang memiliki nilai *accuracy* 68%, *precision* 74%, *recall* 68%, dan *f1-score* 61%.

**Kata kunci:** Aplikasi, Analisis Sentimen, Algoritma Naïve Bayes, Google Play Store, Haji Pintar, *K-Fold Cross Validation*, *Term Frequency - Inverse Document Frequency*





# Sentiment Analysis of Reviews on Haji Pintar Application Using Naïve Bayes Algorithm

Muhammad Rezalutfi

## ABSTRACT

Haji Pintar is an application used by Haji pilgrims in Indonesia to obtain information related to the implementation of Hajj and assist pilgrims during the Hajj pilgrimage. Various user reviews of the Haji Pintar application on the Google Play Store contain complaints and suggestions about the application. Therefore, a research study was conducted to analyze the positive, neutral, and negative sentiments of user reviews of the Haji Pintar application on the Google Play Store. This study employed the Text Preprocessing method, which includes Data Cleaning, Case Folding, Tokenization, Stopword Removal, and Stemming, followed by Term Frequency - Inverse Document Frequency weighting. The Naïve Bayes algorithm was then used as the classification model with data splitting and K-Fold Cross Validation. The best test result was obtained by the Naïve Bayes classification model using an 80:20 Data Split, which achieved an accuracy of 68%, precision of 74%, recall of 68%, and an f1-score of 61%.

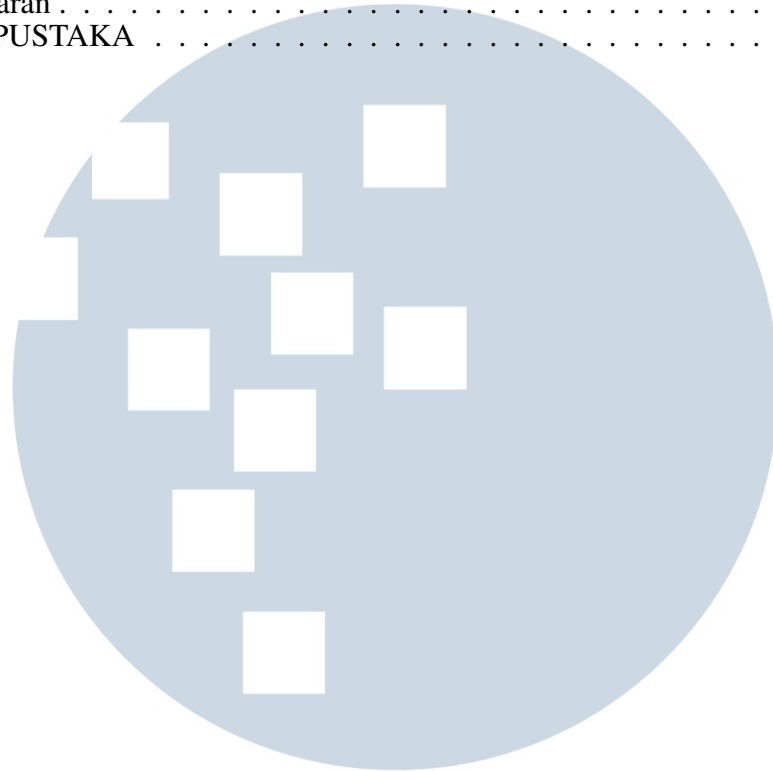
**Keywords:** Application, Google Play Store, Haji Pintar, K-Fold Cross Validation, Naïve Bayes Algorithm, Sentiment Analysis, Term Frequency - Inverse Document Frequency



## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PENGESAHAN . . . . .	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iv
HALAMAN PERSEMBAHAN/MOTO . . . . .	v
KATA PENGANTAR . . . . .	vi
ABSTRAK . . . . .	vii
ABSTRACT . . . . .	viii
DAFTAR ISI . . . . .	ix
DAFTAR GAMBAR . . . . .	xi
DAFTAR TABEL . . . . .	xii
DAFTAR LAMPIRAN . . . . .	xiii
BAB 1 PENDAHULUAN . . . . .	1
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	4
1.3 Batasan Permasalahan . . . . .	4
1.4 Tujuan Penelitian . . . . .	4
1.5 Manfaat Penelitian . . . . .	4
1.6 Sistematika Penulisan . . . . .	5
BAB 2 LANDASAN TEORI . . . . .	7
2.1 Analisis Sentimen . . . . .	7
2.2 Text Preprocessing . . . . .	7
2.3 Term Frequency - Inverse Document Frequency (TF-IDF) . . . . .	8
2.4 Naïve Bayes . . . . .	9
2.5 K-Fold Cross Validation . . . . .	9
2.6 Confusion Matrix . . . . .	10
BAB 3 METODOLOGI PENELITIAN . . . . .	12
3.1 Metodologi Penelitian . . . . .	12
3.2 Perancangan Sistem . . . . .	13
3.2.1 Teknik Pengumpulan Data . . . . .	13
3.2.2 Flowchart Text Preprocessing . . . . .	13
3.2.3 Data Labeling . . . . .	14
3.2.4 Flowchart Term Frequency - Inverse Document Frequency . . . . .	15
3.2.5 Flowchart Naive Bayes . . . . .	16
BAB 4 HASIL DAN DISKUSI . . . . .	18
4.1 Spesifikasi Sistem . . . . .	18
4.2 Teknik Pengumpulan Data . . . . .	18
4.3 Implementasi Sistem . . . . .	19
4.3.1 Potongan Kode Text Preprocessing . . . . .	20
4.3.2 Potongan Kode <i>Data Labeling</i> . . . . .	27
4.3.3 Potongan Kode Pembobotan Term Frequency - Inverse Document Frequency . . . . .	30
4.3.4 Potongan Kode Naïve Bayes dengan Data Split . . . . .	30
4.3.5 Potongan Kode Naïve Bayes dengan K-Fold Cross Validation . . . . .	31
4.4 Uji Coba dan Evaluasi . . . . .	32
4.4.1 Uji Coba dengan Data Split . . . . .	32
4.4.2 Uji Coba dengan K-Fold Cross Validation . . . . .	35
4.4.3 Evaluasi Hasil Pengujian . . . . .	38

BAB 5	SIMPULAN DAN SARAN . . . . .	39
5.1	Simpulan . . . . .	39
5.2	Saran . . . . .	39
DAFTAR PUSTAKA	. . . . .	40



UMMN  
 UNIVERSITAS  
 MULTIMEDIA  
 NUSANTARA

## DAFTAR GAMBAR

Gambar 1.1	Daftar negara dengan kuota haji 2022 terbanyak sumber: [1]	1
Gambar 1.2	Aplikasi Haji Pintar sumber: Google Play Store . . . . .	2
Gambar 2.1	Ilustrasi <i>K-Fold Cross Validation</i> sumber: [2] . . . . .	10
Gambar 3.1	<i>Flowchart Text Preprocessing</i> . . . . .	14
Gambar 3.2	<i>Flowchart</i> pembobotan <i>Flowchart</i> TF-IDF . . . . .	15
Gambar 3.3	<i>Flowchart</i> Naïve Bayes dengan <i>Data Split</i> . . . . .	16
Gambar 3.4	<i>Flowchart</i> Naïve Bayes dengan <i>K-Fold Cross Validation</i> . . . . .	17
Gambar 4.1	Potongan kode <i>Scraping Data</i> . . . . .	19
Gambar 4.2	Potongan kode <i>import library</i> . . . . .	19
Gambar 4.3	Potongan kode <i>import</i> dataset . . . . .	20
Gambar 4.4	Dataset . . . . .	20
Gambar 4.5	Potongan kode <i>data cleaning</i> . . . . .	21
Gambar 4.6	Potongan kode <i>case folding</i> . . . . .	22
Gambar 4.7	Potongan kode <i>tokenization</i> . . . . .	23
Gambar 4.8	Potongan kode <i>stopword removal</i> . . . . .	24
Gambar 4.9	Potongan kode <i>stemming</i> . . . . .	25
Gambar 4.10	Potongan kode penghapusan data kosong . . . . .	26
Gambar 4.11	Hasil <i>text preprocessing</i> . . . . .	26
Gambar 4.12	<i>Wordcloud</i> data ulasan aplikasi Haji Pintar . . . . .	27
Gambar 4.13	Hasil terjemahan ke bahasa Inggris . . . . .	27
Gambar 4.14	Potongan kode <i>data labeling</i> dengan TextBlob . . . . .	28
Gambar 4.15	Hasil <i>data labeling</i> dengan TextBlob . . . . .	28
Gambar 4.16	<i>Wordcloud</i> data ulasan positif . . . . .	29
Gambar 4.17	<i>Wordcloud</i> data ulasan negatif . . . . .	29
Gambar 4.18	<i>Wordcloud</i> data ulasan netral . . . . .	30
Gambar 4.19	Potongan kode TF-IDF . . . . .	30
Gambar 4.20	Potongan kode Naïve Bayes dengan <i>Data Split</i> . . . . .	31
Gambar 4.21	Potongan kode Naïve Bayes dengan <i>k-fold cross validation</i> . . . . .	31
Gambar 4.22	Hasil evaluasi Naïve Bayes dengan <i>data split</i> 50:50 . . . . .	32
Gambar 4.23	Hasil evaluasi Naïve Bayes dengan <i>data split</i> 60:40 . . . . .	33
Gambar 4.24	Hasil evaluasi Naïve Bayes dengan <i>data split</i> 70:30 . . . . .	33
Gambar 4.25	Hasil evaluasi Naïve Bayes dengan <i>data split</i> 80:20 . . . . .	34
Gambar 4.26	Hasil evaluasi Naïve Bayes dengan <i>data split</i> 90:10 . . . . .	34
Gambar 4.27	Hasil evaluasi Naïve Bayes dengan <i>k-fold cross validation</i> k = 6 . . . . .	35
Gambar 4.28	Hasil evaluasi Naïve Bayes dengan <i>k-fold cross validation</i> k = 7 . . . . .	36
Gambar 4.29	Hasil evaluasi Naïve Bayes dengan <i>k-fold cross validation</i> k = 8 . . . . .	36
Gambar 4.30	Hasil evaluasi Naïve Bayes dengan <i>k-fold cross validation</i> k = 9 . . . . .	37
Gambar 4.31	Hasil evaluasi Naïve Bayes dengan <i>k-fold cross validation</i> k = 10 . . . . .	37

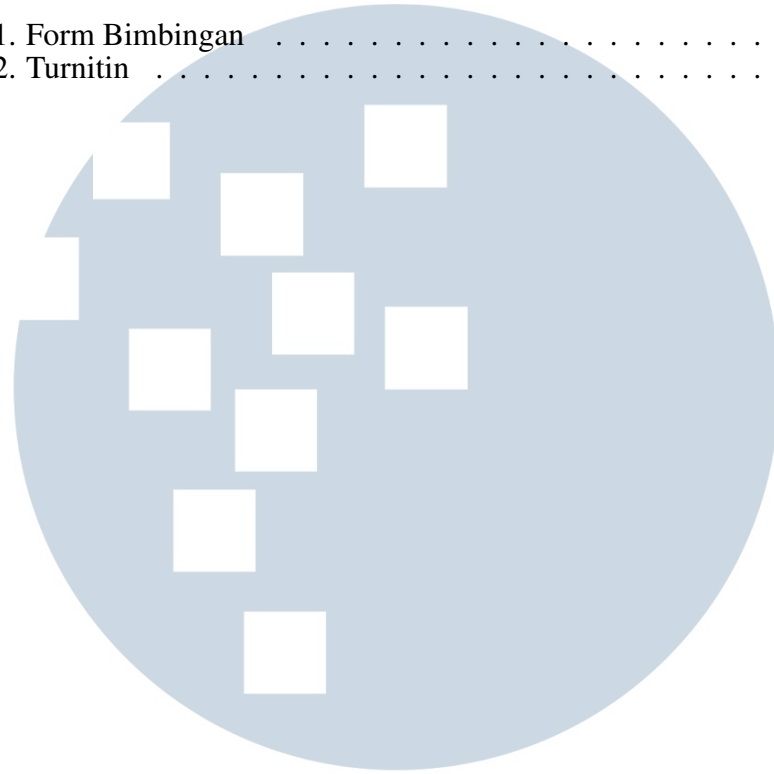
## DAFTAR TABEL

Tabel 2.1	<i>Confusion Matrix</i> . . . . .	10
Tabel 4.1	Hasil <i>data cleaning</i> . . . . .	21
Tabel 4.2	Hasil <i>case folding</i> . . . . .	22
Tabel 4.3	Hasil <i>tokenization</i> . . . . .	23
Tabel 4.4	Hasil <i>stopword removal</i> . . . . .	24
Tabel 4.5	Hasil <i>stemming</i> . . . . .	25
Tabel 4.6	Perbandingan Naive Bayes menggunakan <i>data split</i> . . . . .	38
Tabel 4.7	Perbandingan Naive Bayes menggunakan <i>k-fold cross validation</i> . . . . .	38



## DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan .....	42
Lampiran 2. Turnitin .....	44



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA