



**ANALISIS MINAT CALON MURID DALAM MEMILIH MTS
MAMBAUL ULUM PADA SPMB MENGGUNAKAN
ALGORITMA NAIVE BAYES**

SKRIPSI

disusun oleh:

MOCH JANWAR ABDUL AZIS

NIM. 22166024

**PROGRAM STUDI SARJANA SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS HARKAT NEGERI**

2026



**ANALISIS MINAT CALON MURID DALAM MEMILIH MTS
MAMBAUL ULUM PADA SPMB MENGGUNAKAN
ALGORITMA NAIVE BAYES**

***ANALYSIS OF PROSPECTIVE STUDENTS' INTEREST IN
CHOOSING MTS MAMBAUL ULUM IN SPMB USING THE
NAIVE BAYES ALGORITHM***

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi

disusun oleh :

MOCH JANWAR ABDUL AZIS

NIM. 22166024

**PROGRAM STUDI SARJANA SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS HARKAT NEGERI**

2026

HALAMAN PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama : Moch Janwar Abdul Azis
NIM : 22166024
Jurusan / Program Studi : Sistem Informasi
Jenis Karya : Skripsi

Adalah mahasiswa Program Studi S1 Sistem Informasi, dengan ini saya menyatakan bahwa laporan Skripsi yang berjudul “ANALISIS MINAT CALON MURID DALAM MEMILIH MTS MAMBAUL ULUM PADA SPMB MENGGUNAKAN ALGORITMA NAIVE BAYES”

Merupakan hasil pemikiran dan kerja sama sendiri secara orisinil dan saya susun secara mandiri dan tidak melanggar kode etika hak karya cipta. Pada pelaporan Skripsi ini juga bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik tertentu di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau disebutkan dalam daftar pustaka.

Apabila di kemudian hari ternyata Laporan Skripsi ini terbukti melanggar kode etik karya cipta atau merupakan karya yang dikategorikan mengandung unsur plagiarisme, maka saya bersedia untuk melakukan penelitian baru dan menyusun laporannya sebagai Laporan Skripsi, sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan sesungguhnya

Tegal, 19 Mei 2026



(Moch Janwar Abdul Azis)

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademika Universitas Harkat Negeri Tegal, Saya yang bertanda tangan dibawah ini :

Nama : Moch Janwar Abdul Azis
NIM : 22166024
Jurusan / Program Studi : Sistem Informasi / S1
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Harkat Negeri Tegal **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas Skripsi saya yang berjudul :

ANALISIS MINAT CALON MURID DALAM MEMILIH MTS MAMBAUL ULUM PADA SPMB MENGGUNAKAN ALGORITMA NAIVE BAYES Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti *Noneksklusif* ini Universitas Harkat Negeri Tegal 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 pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Tegal
Pada Tanggal : 30 Juli 2026

Yang menvatakan



(Moch Janwar Abdul Azis)

HALAMAN PERSETUJUAN

ANALISIS MINAT CALON MURID DALAM MEMILIH MTS MAMBAUL ULUM PADA SPMB MENGGUNAKAN ALGORITMA NAIVE BAYES

ANALYSIS OF PROSPECTIVE STUDENTS' INTEREST IN CHOOSING MTS MAMBAUL ULUM IN SPMB USING THE NAIVE BAYES ALGORITHM

yang disusun dan diajukan oleh

MOCH JANWAR ABDUL AZIS

22166024

telah disetujui oleh Dosen Pembimbing SKRIPSI
pada tanggal 14 Juli 2026

Dosen Pembimbing I



R. Bangkit Indarmawan N., M.Kom
NIPY. 08.025.558

Dosen Pembimbing II



Nugroho Adhi Santoso, S.Kom., M.Kom
NIPY. 08.025.560

HALAMAN PENGESAHAN

**ANALISIS MINAT CALON MURID DALAM MEMILIH MTS
MAMBAUL ULUM PADA SPMB MENGGUNAKAN ALGORITMA
NAIVE BAYES**
*ANALYSIS OF PROSPECTIVE STUDENTS' INTEREST IN CHOOSING
MTS MAMBAUL ULUM IN SPMB USING THE NAIVE BAYES
ALGORITHM*

yang disusun dan diajukan oleh

Moch Janwar Abdul Azis

22166024

Telah dipertahankan di depan Dewan Penguji

pada tanggal 30 Juli 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mohammad Humam, S.Kom., M.Kom.

NIPY. 12.002.007

Safar Dwi Kurniawan, M.Kom.

NIPY. 03021487

Nugroho Adhi Santoso, S.Kom., M.Kom.

NIPY. 08.025.560

Skripsi ini telah diterima sebagai salah satu persyaratan

untuk memperoleh gelar Sarjana Komputer

Tanggal 30 Juli 2026

Kepala Program Studi S1 Sistem Informasi



R. Bangkit Indarmawan N. M.Kom

NIPY. 08.025.558

HALAMAN MOTO

"Maka sesungguhnya bersama kesulitan ada kemudahan."

(QS. Al-Insyirah: 6)

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."

(QS. Al-Baqarah: 286)

"Sesungguhnya Allah bersama orang-orang yang sabar."

(QS. Al-Baqarah: 153)

HALAMAN PERSEMBAHAN

Dengan mengucapkan syukur kepada Allah SWT atas segala rahmat dan karunia-Nya, karya ilmiah ini penulis persembahkan kepada:

1. Kedua orang tua tercinta yang senantiasa memberikan doa, dukungan, serta motivasi dalam setiap langkah kehidupan penulis.
2. Keluarga yang selalu memberikan semangat dan dukungan selama proses penyusunan skripsi.
3. Dosen pembimbing dan seluruh dosen yang telah memberikan ilmu, arahan, serta bimbingan kepada penulis.
4. Almamater UNIVERSITAS HARKAT NEGERI yang menjadi tempat penulis menimba ilmu dan mengembangkan kemampuan akademik.

INTISARI

Penelitian ini bertujuan untuk menganalisis tingkat minat murid yang telah memilih MTs Mambaul Ulum berdasarkan faktor lokasi sekolah, kualitas sekolah, fasilitas sekolah, dan biaya pendidikan menggunakan algoritma *Naive Bayes*. Data penelitian diperoleh melalui penyebaran kuesioner secara daring kepada murid MTs Mambaul Ulum yang telah memilih dan bersekolah di MTs Mambaul Ulum pada tahun ajaran 2023, 2024, dan 2025. Populasi dalam penelitian ini berjumlah 200 murid, sedangkan penentuan jumlah sampel menggunakan rumus *Slovin* dengan tingkat kesalahan 10% sehingga diperoleh 67 responden. Data yang diperoleh merupakan persepsi retrospektif responden berdasarkan pengalaman setelah memilih dan bersekolah di MTs Mambaul Ulum. Variabel yang digunakan terdiri atas lokasi sekolah, kualitas sekolah, fasilitas sekolah, biaya pendidikan, dan minat.

Tahapan penelitian terdiri atas pengumpulan data, pemberian skor menggunakan Skala Likert, uji validitas, uji reliabilitas, transformasi data menggunakan metode interval kelas, analisis deskriptif data, pemodelan menggunakan algoritma *Naive Bayes*, serta evaluasi model menggunakan metode *10-Fold Cross Validation* pada perangkat lunak *RapidMiner*. Data hasil kuesioner dikelompokkan ke dalam kategori rendah, sedang, dan tinggi. Variabel lokasi sekolah, kualitas sekolah, fasilitas sekolah, dan biaya pendidikan digunakan sebagai atribut prediktor, sedangkan variabel minat digunakan sebagai label yang diprediksi.

Hasil analisis deskriptif menunjukkan bahwa kategori tinggi mendominasi pada atribut lokasi sekolah sebanyak 42 responden (62,69%), kualitas sekolah sebanyak 41 responden (61,19%), fasilitas sekolah sebanyak 36 responden (53,73%), dan minat sebanyak 39 responden (58,21%). Sementara itu, atribut biaya pendidikan paling banyak berada pada kategori sedang dengan 33 responden (49,25%). Hasil evaluasi menggunakan metode *10-Fold Cross Validation* menunjukkan bahwa algoritma *Naive Bayes* memperoleh *accuracy* rata-rata sebesar $61,67\% \pm 17,33\%$. Nilai *precision* untuk kategori tinggi, sedang, dan rendah masing-masing sebesar 68,18%, 31,25%, dan 85,71%, sedangkan nilai *recall* masing-masing sebesar 76,92%, 25,00%, dan 75,00%. Berdasarkan hasil tersebut, algoritma *Naive Bayes* dapat digunakan untuk mengklasifikasikan tingkat minat murid yang telah memilih MTs Mambaul Ulum, tetapi kemampuan model masih terbatas terutama dalam mengidentifikasi kategori sedang. Hasil penelitian dapat digunakan sebagai informasi pendukung bagi pihak sekolah dalam memahami karakteristik tingkat minat responden yang telah memilih MTs Mambaul Ulum.

Kata kunci: *Data Mining, Naive Bayes, Minat Murid, Klasifikasi, SPMB.*

ABSTRACT

This study aims to analyze the level of interest of students who have chosen MTs Mambaul Ulum based on the factors of school location, school quality, school facilities, and education costs using the *Naive Bayes* algorithm. The research data were obtained through an online questionnaire distributed to students of MTs Mambaul Ulum who had chosen and studied at MTs Mambaul Ulum in the 2023, 2024, and 2025 academic years. The population in this study consisted of 200 students, while the sample size was determined using the Slovin formula with a 10% margin of error, resulting in 67 respondents. The data obtained represent the respondents' retrospective perceptions based on their experiences after choosing and studying at MTs Mambaul Ulum. The variables used consisted of school location, school quality, school facilities, education costs, and interest.

The research stages consisted of data collection, scoring using a *Likert Scale*, validity testing, reliability testing, data transformation using the class interval method, descriptive data analysis, modeling using the *Naive Bayes* algorithm, and model evaluation using the *10-Fold Cross Validation* method with *RapidMiner* software. The questionnaire data were categorized into low, medium, and high categories. School location, school quality, school facilities, and education costs were used as predictor attributes, while interest was used as the label to be predicted. The descriptive analysis results showed that the high category dominated the school location attribute with 42 respondents (62.69%), school quality with 41 respondents (61.19%), school facilities with 36 respondents (53.73%), and interest with 39 respondents (58.21%). Meanwhile, the education cost attribute was predominantly in the medium category, with 33 respondents (49.25%). The evaluation results using the *10-Fold Cross Validation* method showed that the *Naive Bayes* algorithm achieved an average *accuracy* of $61.67\% \pm 17.33\%$. The *precision* values for the high, medium, and low categories were 68.18%, 31.25%, and 85.71%, respectively, while the recall values were 76.92%, 25.00%, and 75.00%, respectively. Based on these results, the *Naive Bayes* algorithm can be used to classify the level of interest of students who have chosen MTs Mambaul Ulum, but the model's performance remains limited, particularly in identifying the medium category. The research results can be used as supporting information for the school in understanding the characteristics of the interest level of respondents who have chosen MTs Mambaul Ulum.

Keywords: *Data Mining, Naive Bayes, Student Interest, Classification, SPMB.*

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT, Tuhan Yang Maha Pengasih dan Maha Penyayang yang telah melimpahkan segala rahmat, hidayah dan inayah-Nya hingga terselesaikannya laporan Skripsi dengan judul “Analisis Minat Calon Murid dalam Memilih MTs Mambaul Ulum pada SPMB Menggunakan Algoritma *Naive Bayes*”.

Skripsi merupakan suatu kewajiban yang harus dilaksanakan untuk memenuhi salah satu syarat kelulusan dalam mencapai derajat Sarjana Sistem Informasi pada Program Studi S1 Sistem Informasi Universitas Harkat Negeri Tegal. Selama melaksanakan penelitian dan kemudian tersusun dalam laporan Skripsi ini, banyak pihak yang telah memberikan bantuan, dukungan dan bimbingan.

Pada kesempatan ini, tidak lupa diucapkan terima kasih yang sebesar besarnya kepada:

1. Bapak Sudirman Said, A.K., M.B.A Selaku Rektor Universitas Harkat Negeri Tegal
2. Bapak Dr. Irendra Radjawali, Ph.D. Selaku Dekan Fakultas Sains dan Teknologi Universitas Harkat Negeri Tegal.
3. Bapak R. Bangkit Indarmawan N. S.Kom., M.Kom. Selaku Kaprodi S1 Sistem Informasi Sekaligus Dosen Pembimbing I.
4. Bapak Nugroho Adhi Santoso, S.Kom., M.Kom. Selaku Dosen Pembimbing II.
5. Kedua orang tua dan keluarga tercinta yang senantiasa memberikan doa, dukungan, motivasi, dan kasih sayang kepada penulis.
6. Pihak MTs Mambaul Ulum yang telah memberikan izin dan membantu dalam penyediaan data penelitian.
7. Teman-teman serta semua pihak yang telah membantu dan memberikan dukungan dalam penyusunan skripsi ini.

Semoga laporan Skripsi ini dapat memberikan sumbangan untuk pengembangan ilmu pengetahuan dan teknologi.

Tegal, 30 Juli 2026

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KEASLIAN.....	ii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN.....	v
HALAMAN MOTO.....	vi
INTISARI	viii
ABSTRACT.....	ix
KATA PENGANTAR.....	x
DAFTAR ISI.....	xii
DAFTAR TABEL	xv
DAFTAR GAMBAR.....	xvi
DAFTAR LAMPIRAN	xvii
DAFTAR LAMBANG DAN SINGKATAN	xviii
DAFTAR ISTILAH	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA	10
2.1 Landasan Teori	10
2.1.1 <i>Data Mining</i>	10
2.1.2 <i>Educational Data Mining</i>	11
2.1.3 Pertimbangan dalam Memilih Sekolah	12
2.1.4 Klasifikasi	14
2.1.5 Algoritma <i>Naive Bayes</i>	15
2.1.6 <i>RapidMiner</i>	16

2.1.7 <i>Cross Validation</i>	17
2.1.8 <i>Confusion Matrix</i>	18
2.2 Penelitian Terdahulu	20
BAB III METODE PENELITIAN	28
3.1 Jenis, Sifat, dan Pendekatan Penelitian.....	28
3.2 Populasi dan Sampel.....	29
3.2.1 Populasi	29
3.2.2 Sampel.....	29
3.3 Teknik pengumpulan data.....	30
3.4 Tempat dan Waktu Penelitian.....	32
3.5 Instrumen penelitian	32
3.5.1 Skala Pengukuran.....	35
3.5.2 Uji Validitas	36
3.5.3 Uji Reliabilitas	38
3.6 Metode Analisis Data.....	39
3.7 Alur Penelitian	43
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	46
4.1 Hasil Penelitian	46
4.1.1 Pengumpulan Data	47
4.1.2 Skoring Kuesioner	48
4.1.3 Uji Validitas	49
4.1.4 Uji Reliabilitas	50
4.1.5 Transformasi Data Menggunakan Metode Interval Kelas	51
4.1.6 Hasil Transformasi Data	52
4.1.7 Analisis Deskriptif Data.....	55
4.1.8 Pemodelan Menggunakan Algoritma <i>Naive Bayes</i>	57
4.1.9 Hasil Evaluasi Model	59
4.2 Pembahasan	62
BAB V PENUTUP.....	66
5.1 Kesimpulan	66
5.2 Saran	67

DAFTAR PUSTAKA.....	69
LAMPIRAN.....	72

DAFTAR TABEL

Tabel 2.1 Matriks literatur review dan posisi penelitian.....	20
Tabel 3.1 Operasional atribut penelitian	33
Tabel 3.2 Skala Likert.....	36
Tabel 4.1 Skor Skala Likert pernyataan positif	48
Tabel 4.2 Hasil uji validitas	49
Tabel 4.3 Hasil uji reliabilitas	51
Tabel 4.4 Hasil perhitungan interval kelas.....	52
Tabel 4.5 Hasil transformasi data	53
Tabel 4.6 Distribusi kategori atribut penelitian	55

DAFTAR GAMBAR

Gambar 3.1 Alur penelitian.....	45
Gambar 4.1 Proses pemodelan <i>Naive Bayes</i> menggunakan <i>RapidMiner</i>	58
Gambar 4.2 Proses <i>training</i> dan <i>testing</i> pada operator <i>Cross Validation</i>	59
Gambar 4.3 Hasil evaluasi model menggunakan algoritma <i>Naive Bayes</i>	60

DAFTAR LAMPIRAN

Lampiran A-1 Kuesioner Penelitian	73
Lampiran A-2 Data Hasil Kuesioner 67 Responden.....	80
Lampiran A-3 Hasil Uji Reliabilitas	85
Lampiran A-4 Formulir Perubahan Judul Skripsi.....	86
Lampiran A-5 Surat Pengantar Penelitian	87
Lampiran A-6 Surat Keterangan	88
Lampiran A-7 Jadwal Kegiatan	89
Lampiran A-8 Histori Bimbingan Skripsi.....	90
Lampiran A-9 Curriculum Vitae.....	94
Lampiran A-10 Formulir publikasi jurnal skripsi.....	96
Lampiran A-11 Letter of Acceptance (LOA)	97
Lampiran A-12 Bukti Cek Plagiarisme.....	98
Lampiran A-13 Dokumentasi kegiatan.....	99

DAFTAR LAMBANG DAN SINGKATAN

SPMB	Sistem Penerimaan Murid Baru
MTs	Madrasah Tsanawiyah
X1	Variabel Lokasi Sekolah
X2	Variabel Kualitas Sekolah
X3	Variabel Fasilitas Sekolah
X4	Variabel Biaya Pendidikan
Y	Variabel Minat
$P(H X)$	Probabilitas hipotesis H berdasarkan kondisi X
$P(X H)$	Probabilitas munculnya X jika H terjadi
$P(H)$	Probabilitas awal hipotesis H
$P(X)$	Probabilitas munculnya X
H	Hipotesis atau kelas yang diprediksi
X	Data atau atribut yang digunakan sebagai dasar prediksi
r_{xy}	Koefisien korelasi <i>Pearson Product Moment</i>
N	Jumlah responden
ΣX	Jumlah skor setiap butir pernyataan
ΣY	Jumlah skor total pada konstruk yang sama
ΣXY	Jumlah hasil perkalian skor setiap butir pernyataan dengan skor total pada konstruk yang sama
ΣX^2	Jumlah kuadrat skor setiap butir pernyataan
ΣY^2	Jumlah kuadrat skor total pada konstruk yang sama
α	Koefisien reliabilitas <i>Cronbach's Alpha</i>
k	Jumlah butir pernyataan dalam konstruk
ΣS_i^2	Jumlah varians setiap butir pernyataan dalam konstruk
S_t^2	Varians total skor konstruk
C	Interval kelas
X_n	Skor maksimum
X_i	Skor minimum

K	Jumlah kategori
P	Persentase responden
f	Jumlah responden pada suatu kategori

DAFTAR ISTILAH

<i>Data Mining</i>	Proses pengolahan data untuk menemukan pola atau informasi yang bermanfaat.
Klasifikasi	Teknik untuk mengelompokkan data ke dalam kelas atau kategori tertentu.
<i>Naive Bayes</i>	Algoritma klasifikasi berbasis probabilitas yang menggunakan <i>Teorema Bayes</i> .
<i>Teorema Bayes</i>	Teorema probabilitas yang menjadi dasar perhitungan algoritma <i>Naive Bayes</i> .
Probabilitas	Nilai yang menunjukkan kemungkinan suatu kejadian atau data.
Atribut	Karakteristik atau variabel yang digunakan sebagai dasar klasifikasi.
Prediktor	Atribut yang digunakan untuk memprediksi kelas atau label.
Label	Kelas atau kategori yang menjadi target prediksi.
<i>Dataset</i>	Kumpulan data yang digunakan dalam proses analisis dan klasifikasi.
Transformasi Data	Proses mengubah data ke dalam bentuk yang sesuai untuk analisis.
Skala Likert	Skala pengukuran tingkat persetujuan responden terhadap pernyataan.
Interval Kelas	Rentang nilai yang digunakan untuk membentuk kategori data.
Analisis Deskriptif Data	Analisis yang digunakan untuk menggambarkan distribusi data berdasarkan jumlah dan persentase.
Kuesioner	Instrumen yang digunakan untuk memperoleh data dari responden.
Responden	Pihak yang memberikan jawaban pada kuesioner.

Persepsi Retrospektif	Persepsi responden berdasarkan pengalaman setelah memilih dan bersekolah di MTs Mambaul Ulum.
Uji Validitas	Pengujian untuk mengetahui ketepatan instrumen dalam mengukur variabel penelitian.
Uji Reliabilitas	Pengujian untuk mengetahui konsistensi instrumen penelitian.
<i>Pearson Product Moment</i>	Metode korelasi yang digunakan untuk menguji validitas butir pernyataan.
<i>Cronbach's Alpha</i>	Metode yang digunakan untuk mengukur reliabilitas instrumen.
<i>10-Fold Cross Validation</i>	Metode pengujian dengan membagi data menjadi sepuluh bagian untuk proses pelatihan dan pengujian secara bergantian.
<i>Confusion Matrix</i>	Metode evaluasi untuk membandingkan hasil prediksi dengan kelas sebenarnya.
<i>Accuracy</i>	Ukuran tingkat ketepatan model dalam mengklasifikasikan data.
<i>Precision</i>	Ukuran ketepatan model dalam memprediksi suatu kelas.
<i>Recall</i>	Ukuran kemampuan model dalam mengenali data pada suatu kelas.
<i>RapidMiner</i>	Perangkat lunak yang digunakan untuk pemodelan dan evaluasi <i>Naive Bayes</i> .
Minat	Tingkat ketertarikan responden dalam memilih MTs Mambaul Ulum.
Lokasi Sekolah	Atribut yang berkaitan dengan jangkauan, posisi, dan akses menuju sekolah.
Kualitas Sekolah	Atribut yang menggambarkan kualitas pembelajaran dan layanan pendidikan.

Fasilitas Sekolah

Sarana dan prasarana yang mendukung kegiatan pembelajaran.

Biaya Pendidikan

Biaya yang berkaitan dengan kebutuhan dalam memperoleh layanan pendidikan.