

**PENERAPAN ALGORITMA *K-MEANS* DAN *NAÏVE BAYES*
UNTUK MENGANALISIS FAKTOR PENENTU PRESTASI
AKADEMIK BAHASA INGGRIS SISWA SMA SWASTA
BHAYANGKARI 2 RANTAU PRAPAT**

SKRIPSI

Untuk Memenuhi Persyaratan Memperoleh Gelar Sarjana (S1)
Program Studi Sistem Informasi Fakultas Sains dan Teknologi
Universitas Labuhanbatu



OLEH:

ANGGI AUDYA

2209100013

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAIN DAN TEKNOLOGI
UNIVERSITAS LABUHANBATU
RANTAU PRAPAT**

2026

LEMBAR PERSETUJUAN SKRIPSI

JUDUL SKRIPSI : PENERAPAN ALGORITMA K-MEANS DAN
NAIVE BAYES UNTUK MENGANALISIS
FAKTOR PENENTU PRESTASI AKADEMIK
BAHASA INGGRIS SISWA SMA SWASTA
BHAYANGKARI 2 RANTAU PRAPAT

NAMA : ANGGI AUDYA
NPM : 2209100013
PRODI : SISTEM INFORMASI

Disetujui pada tanggal : 09 April 2026

Pembimbing I



Syaiful Zuhri Harahap, S.Kom., M.Kom
NIDN.0113129103

Pembimbing II



Asriani Hasibuan, S.Pd., M.Hum
NIDN. 0116078802

LEMBAR PENGESAHAN SKRIPSI

JUDUL SKRIPSI : PENERAPAN ALGORITMA K-MEANS DAN
NAIVE BAYES UNTUK MENGANALISIS
FAKTOR PENENTU PRESTASI AKADEMIK
BAHASA INGGRIS SISWA SMA SWASTA
BHAYANGKARI 2 RANTAU PRAPAT

NAMA : ANGGI AUDYA
NPM : 2209100013
PRODI : SISTEM INFORMASI

Telah Diuji dan Dinyatakan Lulus Dalam Uji Sarjana Pada
Tanggal 09 April 2026

TIM PENGUJI

Pembimbing I (Ketua)

Nama : Syaiful Zuhri Harahap, S.Kom., M.Kom ()
NIDN : 0113129103

Tanda Tangan



Pembimbing II (Anggota)

Nama : Asriani Hasibuan, S.Pd., M.Hum ()
NIDN : 0116078802



Penguji (Anggota)

Nama : Budianto Bangun, S.Sos., M.Kom ()
NIDN : 0124047003

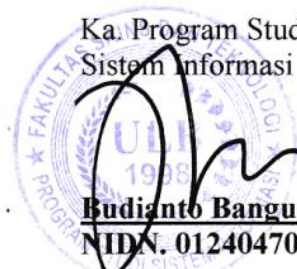


Rantauprapat, 09 April 2026



Dekan
Fakultas Sains dan Teknologi

Assoc. Prof. Dr. Iwan Purnama, S.Kom., M.Kom
NIDN. 0112029202



Ka. Program Studi
Sistem Informasi

Budianto Bangun, S.Sos., M.Kom
NIDN. 0124047003

PERNYATAAN

JUDUL SKRIPSI : PENERAPAN ALGORITMA K-MEANS DAN
NAIVE BAYES UNTUK MENGANALISIS
FAKTOR PENENTU PRESTASI AKADEMIK
BAHASA INGGRIS SISWA SMA SWASTA
BHAYANGKARI 2 RANTAU PRAPAT
NAMA : ANGGI AUDYA
NPM : 2209100013
PRODI : SISTEM INFORMASI

Dengan ini penulis menyatakan bahwa skripsi ini disusun sebagai syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Labuhanbatu adalah hasil karya tulis penulis sendiri. Semua kutipan maupun rujukan dalam penulisan skripsi ini telah penulis cantumkan sumbernya dengan benar sesuai dengan ketentuan yang berlaku. Jika ada kemudian hari ternyata ditemukan seluruh atau sebagian skripsi ini bukan hasil karya penulis atau plagiat, penulis bersedia menerima sanksi pencabutan gelar akademik yang disandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Rantauprapat, 09 April 2026

Yang Membuat Pernyataan ,



ANGGI AUDYA

2209100013

ABSTRAK

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi prestasi akademik Bahasa Inggris siswa di SMA Swasta Bhayangkari 2 Rantau Prapat dengan menggunakan kombinasi algoritma K-Means dan Naïve Bayes. Data yang digunakan meliputi nilai Bahasa Inggris, kehadiran, keikutsertaan dalam kursus, dan minat belajar siswa. Algoritma K-Means digunakan untuk mengelompokkan siswa ke dalam tiga kategori prestasi, yaitu rendah, sedang, dan tinggi berdasarkan kemiripan karakteristik data, sedangkan Naïve Bayes digunakan untuk mengklasifikasikan serta menentukan faktor dominan yang memengaruhi prestasi akademik. Hasil penelitian menunjukkan bahwa kombinasi kedua algoritma mampu memberikan analisis yang komprehensif, di mana K-Means menghasilkan pengelompokan siswa secara objektif dan Naïve Bayes mengidentifikasi faktor-faktor yang berpengaruh. Faktor seperti minat belajar, kehadiran, dan keikutsertaan dalam kursus Bahasa Inggris terbukti berkontribusi terhadap tingkat prestasi siswa. Dengan demikian, penelitian ini dapat membantu pihak sekolah dalam mengambil keputusan yang lebih tepat dan berbasis data untuk meningkatkan kualitas pembelajaran Bahasa Inggris.

Kata kunci: Data Mining, *K-Means*, *Naïve Bayes*, Prestasi Akademik, Bahasa Inggris

ABSTRACT

This study aims to analyze the factors influencing students' academic achievement in English at SMA Swasta Bhayangkari 2 Rantau Prapat using a combination of K-Means and Naïve Bayes algorithms. The data used include English scores, attendance, participation in English courses, and students' learning interest. The K-Means algorithm is applied to cluster students into three achievement categories-low, medium, and high-based on data similarity, while the Naïve Bayes algorithm is used to classify and determine the dominant factors affecting academic performance. The results show that the combination of both algorithms provides a comprehensive analysis, where K-Means produces objective student grouping and Naïve Bayes identifies influential factors. Factors such as learning interest, attendance, and participation in English courses are proven to contribute to students' performance levels. Therefore, this study can assist schools in making more accurate data-driven decisions to improve the quality of English learning.


Keywords: Data Mining, *K-Means*, *Naïve Bayes*, *Academic Achievement*, *English*

KATA PENGANTAR

Dengan menyebut nama Allah SWT Yang Maha Pengasih lagi Maha Penyayang, penulis mengucapkan puja dan puji syukur ke hadirat-Nya sehingga proposal penelitian berjudul “Penerapan Algoritma K-Means dan Naive Bayes untuk Menganalisis Faktor Penentu Prestasi Akademik Bahasa Inggris Siswa SMA Swasta Bhayangkari 2 Rantau Prapat” ini dengan baik dan tepat waktu. Penulis menyampaikan terima kasih kepada seluruh pihak yang telah memberikan dukungan, khususnya kepada:

1. Alm. Bapak Dr. H. Amarullah Nasution, MBA selaku pendiri Yayasan Universitas Labuhanbatu
2. Bapak Halomoan Nasution, S.H., M.H selaku ketua yayasan Universitas Labuhanbatu.
3. Bapak Rektor Universitas Labuhanbatu, Bapak Assoc, Prof. Ade Parlaungan Nasution, Ph.D
4. Bapak Assoc. Prof. Dr. Iwan Purnama, S.Kom., M.Kom selaku Dekan Fakultas Sains dan Teknologi.
5. Bapak Budianto Bangun, S.Sos, M.Kom selaku Kepala program studi Sistem Informasi dan Selaku Dosen Penguji.
6. Bapak Syaiful Zuhri Harahap, S.Kom., M.Kom selaku Dosen Pembimbing I.
7. Ibu Asriani Hasibuan, S.Pd., M.Hum selaku Dosen Pembimbing II.

Rantau Prapat, April 2026


Anggi Audya

DAFTAR ISI

LEMBAR PERSETUJUAN SKRIPSI.....	i
LEMBAR PENGESAHAN SKRIPSI.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat Penelitian	4
1.4.1 Tujuan Penelitian	4
1.4.2 Manfaat Penelitian	5
1.5 Tinjauan Umum Objek Penelitian.....	6
1.5.1 Struktur Organisasi	6
1.6 Sistematika Penulisan.....	8
BAB II LANDASAN TEORI	10
2.1. <i>Knowledge Discovery in Database</i>	10
2.2. Data Mining	12

2.3. Pengertian Prestasi Akademik.....	13
2.3.1 Faktor yang Mempengaruhi Prestasi Akademik <i>Internal</i> dan <i>Eksternal</i>	15
2.4. Algoritma <i>K-means</i>	16
2.4.1. <i>Clustering</i>	17
2.4.2. Rumus Umum Algoritma K-Means	18
2.5. Algoritma <i>Naive Bayes</i>	19
2.5.1. Rumus Utama <i>Naive Bayes</i>	20
2.6. Penerapan <i>Naive Bayes</i> dan <i>K-means</i> dalam Analisis Faktor Penentu Prestasi Akademik.....	22
2.7. Pengertian Siswa	25
2.8. Prestasi Akademik Bahasa Inggris.....	26
2.9. Alat Bantu Pemrograman untuk Implementasi.....	27
2.9.1. Microsoft Excel.....	28
2.9.2 RapidMiner	29
2.10. Penelitian Terdahulu.....	30
BAB III METODOLOGI PENELITIAN	34
3.1. Kerangka Kerja Penelitian	34
3.2. Lokasi dan Waktu Penelitian.....	38
3.3. Populasi dan Sampel	38
3.4. Variabel Penelitian	39
3.5 Langkah – Langkah Penerapan Algoritma K-Means.....	39
3.6 .Penerapan Algoritma <i>Naive Bayes</i>	55

3.6.1 Evaluasi Model	63
BAB IV HASIL DAN PEMBAHASAN.....	67
4.1. Analisis Data	67
4.2. Pengumpulan Data	68
4.3. Penerapan Algoritma K-Means Dengan RapidMiner	71
4.4. Hasil Penerapan Algoritma K - Means Dengan RapidMiner.....	81
4.5. Penerapan Algoritma <i>Naïve Bayes</i> Dengan RapidMiner	84
4.6. Hasil Penerapan Algoritma <i>Naïve Bayes</i> Dengan RapidMiner	88
4.7. Hasil Pengelompokan K-Means dan Klasifikasi <i>Naïve Bayes</i>	89
4.8. Hasil Analisis Faktor yang Mempengaruhi Prestasi Siswa.....	89
BAB V KESIMPULAN DAN SARAN.....	92
5.1 Kesimpulan	92
5.2 Saran	93
DAFTAR PUSTAKA	94

DAFTAR TABEL

Tabel 2.1	Penelitian Terdahulu.....	31
Tabel 3.1	Data Prestasi Akademik Siswa	40
Tabel 3.2	Data Transformasi Prestasi Akademik Siswa.....	42
Tabel 3.3	Centeroid Awal	44
Tabel 3.4	Hasil Perhitungan Iterasi Ke – 1	45
Tabel 3.5	Nilai Centeroid Baru Iterasi Ke – 2.....	46
Tabel 3.6	Hasil Perhitungan Iterasi Ke- 2	47
Tabel 3.7	Nilai Centeroid Baru Iterasi Ke – 3.....	48
Tabel 3.8	Hasil Perhitungan Iterasi Ke – 3	49
Tabel 3.9	Nilai Centeroid Baru Iterasi Ke- 4	50
Tabel 3.10.	Hasil Perhitungan Iterasi Ke – 4	51
Tabel 3.11	Hasil Cluster 1 (Tinggi).....	53
Tabel 3.12	Hasil Cluster 2 (Sedang)	53
Tabel 3.13	Hasil Cluster 3 (Rendah).....	54
Tabel 3.14.	Data Atribut Siswa untuk Klasifikasi Prestasi Akademik	55
Tabel 3.15.	Data Uji Siswa untuk Klasifikasi Prestasi Akademik	57
Tabel 3.16	Probabilitas Prior Kelas.....	58
Tabel 3.17	Probabilitas Likelihood Setiap Fitur Kelas Prestasi Akademik	59
Tabel 3.18	Hasil Perhitungan Probabilitas Posterior pada 10 Data Testing	61
Tabel 3.19	Perbandingan Label Asli dan Label Prediksi Hasil Klasifikasi Naïve Bayes pada Data Testing (Data Uji)	64
Tabel 3.20	Confusion Matrix Hasil Klasifikasi Prestasi Akademik Siswa	65

Tabel 3.21 Hasil Evaluasi Kinerja Model Naive Bayes	65
Tabel 4.1 Data siswa Semester Ganjil Siswa Kelas X	68
Tabel 4.2 Hasil Cluster_0 (Tinggi).....	81
Tabel 4.3 Hasil Cluster_1 (Sedang)	82
Tabel 4.4 Hasil Cluster_2 (Rendah).....	83

DAFTAR GAMBAR

Gambar 1.1	Struktur Organisasi SMA Swasta BhaSyangkari 2 Rantau Prapat.....	7
Gambar 2.1	Tahapan Dalam <i>Knowledge Discovery in Database</i>	10
Gambar 2.1	Logo Excel.....	28
Gambar 2.2	Logo RapidMiner	29
Gambar 3.1	Kerangka Kerja Penelitian.....	34
Gambar 4.1.	RapidMiner.....	72
Gambar 4.2.	Tampilan Untuk Memulai Lembar Kerja Baru Pada RapidMiner.....	72
Gambar 4.3.	Tampilan Kosong Halaman Lembar Kerja Baru Pada RapidMiner.....	73
Gambar 4.4.	Proses Menginput Operator K – Means Ke Dalam Lembar.....	73
Gambar 4.5.	Proses Memilih Lokasi Penyimpanan Data.....	74
Gambar 4.6.	Proses Memilih Data Yang Akan Di Input	74
Gambar 4.7.	Proses Memilih Label Dari Dataset.....	75
Gambar 4.8.	Proses Menginput Operator K – Means Ke Dalam Lembar Kerja.....	75
Gambar 4.9.	Proses Menentukan Jumlah Cluster.....	76
Gambar 4.10.	Proses Menyambungkan Operator	76
Gambar 4.11.	Hasil Clustering Data	77
Gambar 4.12.	Hasil dari Cluster Model	77
Gambar 4.13.	Hasil Pengelompokan Cluster_0	78
Gambar 4.14.	Hasil Pengelompokan Cluster_1	79

Gambar 4.15 Hasil Pengelompokan Cluster_2	80
Gambar 4.16. Operator Excel.....	85
Gambar 4.17. Impor Data.....	85
Gambar 4.18. Set Role Variabel Targe	86
Gambar 4.19. <i>Design Process</i>	87
Gambar 4.20. <i>Design Process</i>	87
Gambar 4.21. <i>Design Process</i>	88