

**PENERAPAN METODE KLASIFIKASI *NAÏVE BAYES* DAN
K-NEAREST NEIGHBORS UNTUK APLIKASI PREDIKSI
HASIL SELEKSI PENERIMAAN PESERTA DIDIK BARU
DI SMA NEGERI 17 PALEMBANG**



TESIS

**ANANG PURNOMO KURNIAWAN
ENTERPRISE IT INFRASTRUCTURE
212420007**

**PROGRAM STUDI TEKNIK INFORMATIKA – S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
2025**

**PENERAPAN METODE KLASIFIKASI *NAÏVE BAYES* DAN
K-NEAREST NEIGHBORS UNTUK APLIKASI PREDIKSI
HASIL SELEKSI PENERIMAAN PESERTA DIDIK BARU
DI SMA NEGERI 17 PALEMBANG**



**Tesis ini diajukan sebagai salah satu syarat
untuk memperoleh gelar**

MAGISTER KOMPUTER

**ANANG PURNOMO KURNIAWAN
ENTERPRISE IT INFRASTRUCTURE
212420007**

**PROGRAM STUDI TEKNIK INFORMATIKA – S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
2025**

Halaman Pengesahan Pembimbing Tesis

Judul Tesis: PENERAPAN METODE KLASIFIKASI NAÏVE BAYES DAN K-NEAREST NEIGHBORS UNTUK APLIKASI PREDIKSI HASIL SELEKSI PENERIMAAN PESERTA DIDIK BARU DI SMA NEGERI 17 PALEMBANG

Oleh ANANG PURNOMO KURNIAWAN, NIM 212420007, Tesis ini telah disetujui dan disahkan oleh Pembimbing Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE IT INFRASTRUCTURE, Program Pascasarjana Universitas Bina Darma pada 1 September 2025 dan telah dinyatakan LULUS.

Palembang, 1 September 2025
Mengetahui,
Program Studi Teknik Informatika – S2
Universitas Bina Darma
Ketua,



.....
Dr. Usman Ependi, S.Kom., M.Kom.

Pembimbing,



.....
Prof. Dr. Edi Surya Negara, M.Kom

Halaman Pengesahan Penguji Tesis

Judul Tesis: PENERAPAN METODE KLASIFIKASI NAÏVE BAYES DAN K-NEAREST NEIGHBORS UNTUK APLIKASI PREDIKSI HASIL SELEKSI PENERIMAAN PESERTA DIDIK BARU DI SMA NEGERI 17 PALEMBANG

Oleh ANANG PURNOMO KURNIAWAN , NIM 212420007, Tesis ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE IT INFRASTRUCTURE, Program Pascasarjana Universitas Bina Darma pada 1 September 2025 dan telah dinyatakan LULUS.

Palembang, 1 September 2025
Mengetahui,
Program Pascasarjana
Universitas Bina Darma
Direktur,



PROGRAM PASCASARJANA

Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Penguji I,



Prof. Dr. Edi Surya Negara, M.Kom

Penguji II,



M. Izman Herdjansyah, M.M., Ph.D

Penguji III,



Dr. Tata Sutabri, S.Kom., M.M.S.I.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : ANANG PURNOMO KURNIAWAN

NIM : 212420007

Dengan ini menyatakan bahwa:

1. Karya tulis Saya (Tesis) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik baik (Magister) di Universitas Bina Darma;
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing;
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar pustaka;
4. Karena yakin dengan keaslian karya tulis ini, Saya menyatakan bersedia Tesis yang Saya hasilkan di unggah ke internet;
5. Surat Pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 1 September 2025
Yang Membuat Pernyataan,



ANANG PURNOMO KURNIAWAN
NIM: 212420007

ABSTRAK

Meningkatnya ketersediaan data pendidikan telah menciptakan peluang untuk menerapkan teknik *machine learning* (ML) untuk memprediksi hasil penerimaan peserta didik baru, sehingga memungkinkan intervensi tepat waktu untuk meningkatkan kinerja dan retensi akademik. Studi ini menyelidiki kemampuan prediktif dari dua algoritma klasifikasi yang banyak digunakan yaitu K-Nearest Neighbors (KNN) dan Naïve Bayes dalam mengklasifikasikan hasil kelulusan PPDB di SMA Negeri 17 Palembang. Data yang dikumpulkan mencakup periode tiga tahun dari tahun ajaran 2022/2023 hingga tahun ajaran 2025/2026, tidak termasuk tahun ajaran 2024/2025. Kumpulan data tersebut terdiri dari atribut akademik dan demografis, yang menyediakan representasi multifaset dari profil siswa. Kinerja model dinilai menggunakan metrik akurasi, presisi, dan ingatan yang diperoleh dari analisis matriks prediksi. Model KNN mencapai akurasi 96,48%, dengan nilai presisi dan ingatan khusus kelas melebihi 95%, yang menunjukkan keandalan prediktif yang kuat tetapi dengan kesalahan klasifikasi yang kecil. Sebaliknya, pengklasifikasi Naïve Bayes menunjukkan kinerja prediktif yang sempurna, mencapai akurasi, presisi, dan recall 100%, yang mencerminkan kecocokan persis antara hasil prediksi dan hasil aktual. Hasil ini menggarisbawahi kesesuaian Naïve Bayes dalam konteks di mana asumsi independensi fitur selaras dengan properti statistik himpunan data, sekaligus menyoroti ketahanan KNN dalam menangani tugas klasifikasi non-parametrik. Temuan ini memiliki implikasi praktis bagi institusi pendidikan yang ingin menerapkan sistem peringatan dini dan kerangka kerja pengambilan keputusan berbasis data. Dengan mengintegrasikan model prediktif ini ke dalam sistem informasi di sekolah atau Dinas Pendidikan, maka dapat secara proaktif mengidentifikasi permasalahan PPDB, menyesuaikan strategi intervensi, dan mengoptimalkan alokasi sumber daya. Penelitian di masa mendatang sebaiknya berfokus pada validasi lintas institusi, rekayasa fitur yang diperluas, dan pendekatan *hybrid modeling* untuk meningkatkan generalisasi dan adaptasi di berbagai lingkungan pendidikan.

Kata kunci: *educational data mining*, Naïve Bayes, K-Nearest Neighbors, analitik akademik

ABSTRACT

The growing availability of educational data has created opportunities to apply machine learning (ML) techniques for predicting new student admissions, thereby enabling timely interventions to improve academic performance and retention. This study investigates the predictive capabilities of two widely used classification algorithms—k-Nearest Neighbors (KNN) and Naïve Bayes—in classifying high school graduation outcomes. Data were collected from SMA Negeri 17 Palembang, covering a three-year period from the 2022/2023 to the 2025/2026 academic years, excluding 2024/2025. The dataset comprised both academic and demographic attributes, providing a multifaceted representation of student profiles. Model performance was assessed using accuracy, precision, and recall metrics derived from confusion matrix analysis. The KNN model achieved an accuracy of 96.48%, with class-specific precision and recall values exceeding 95%, indicating strong predictive reliability but with minor misclassifications. In contrast, the Naïve Bayes classifier demonstrated perfect predictive performance, attaining 100% accuracy, precision, and recall, reflecting an exact match between predicted and actual outcomes. These results underscore the suitability of Naïve Bayes in contexts where feature independence assumptions align with the dataset's statistical properties, while also highlighting KNN's robustness in handling non-parametric classification tasks. The findings have practical implications for educational institutions seeking to implement early-warning systems and data-driven decision-making frameworks. By integrating these predictive models into institutional information systems, administrators can proactively identify at-risk students, tailor intervention strategies, and optimize resource allocation. Future research should focus on cross-institutional validation, expanded feature engineering, and hybrid modeling approaches to enhance generalizability and adaptability across diverse educational settings.

Keywords: *educational data mining, Naïve Bayes, k-Nearest Neighbors, academic analytics*

MOTTO DAN PERSEMBAHAN

"Just like a water flow"

(seperti air yang mengalir mengikuti medan, melewati rintangan, atau menunggu dengan sabar dan mengumpulkan kekuatan di balik hambatan)

I dedicate this thesis to my family, my colleagues at Dinas Pendidikan Provinsi Sumatera Selatan, my classmates in the Master of Civil Engineering program, and everyone who supported me in completing this work. Their unwavering encouragement, patience, and belief in my potential have been my greatest source of strength. May this work serve as a small tribute to their endless love and guidance.

KATA PENGANTAR

Dengan mengucapkan puji syukur kehadiran Allah SWT atas segala rahmat, nikmat kebahagiaan, serta seluruh anugerah dalam bentuk apapun yang telah dilimpahkan kepada seluruh hamba-hamba-Nya, penulis dapat menyelesaikan Tesis yang berjudul **“Penerapan Metode Klasifikasi *Naïve Bayes* dan *K-Nearest Neighbors* untuk Aplikasi Prediksi Hasil Seleksi Penerimaan Peserta Didik Baru di SMA Negeri 17 Palembang”**.

Dengan selesainya Tesis ini, penulis ingin mengucapkan terimakasih kepada pihak-pihak yang telah banyak membantu dalam penyelesaian Tesis ini. Dalam kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. Orang tua yang telah memberikan dukungan berupa doa, bimbingan, dan materi, istri (Morena Suci Basmark, S.I.Kom, M.M) dan anak-anak tercinta Kheisya, Keano dan Keandra yang merupakan 3k Family serta keluarga besar SBC Family dan H. Jinal Family yang telah memberikan semangat dan doa untuk menyelesaikan tesis ini.
2. Prof. Dr. Ir. Achmad Syarifudin, M.Sc. Selaku Direktur Program Pascasarjana
3. Dr. Usman Ependi, M.Kom. Selaku Ketua Program Studi Teknik Informatika – S2 Program Pascasarjana Universitas Bina Darma
4. Bapak Prof. Dr. Edi Surya Negara, M.Kom selaku dosen pembimbing dalam penyusunan tesis ini beserta Bapak M. Izman Herdiansyah, M.M., Ph.D dan Bapak Dr. Tata Sutabri S.Kom, M.M.S.I. selaku dosen penguji.
5. Bapak/Ibu dosen yang telah memberikan ilmu selama dalam masa perkuliahan.

6. Ibu Mondyaboni, S.E., S.Kom., M.Si, M.Pd selaku Plt. Kepala Dinas Pendidikan Provinsi Sumatera Selatan atas bimbingan dan teladan diberikan dalam melanjutkan studi ke jenjang berikutnya.
7. Ibu Dra. Poniye, M.Pd selaku Kepala Bidang SMA di Dinas Pendidikan Provinsi Sumatera Selatan atas beserta rekan-rekan Dinas Pendidikan Provinsi Sumatera Selatan.
8. Ibu Dra. Hj. Purwastuti Kusumastuti, M.M selaku Kepala SMA Negeri 17 Palembang beserta jajarannya sebagai tempat penelitian, yang telah membantu melengkapi data-data dan ujicoba dalam penelitian ini.
9. Teman-teman serta sahabat-sahabat penulis yang selalu memberikan semangat kepada penulis dalam menyelesaikan Thesis ini.

Semoga tesis ini dapat memberikan manfaat yang sebesar-besarnya pada kita semua. Sebagai manusia biasa yang mempunyai keterbatasan dan kekurangan, maka penulis menyadari bahwa tesis ini masih banyak kekurangan-kekurangan baik dalam teknik penulisan, pembahasan dan penyajian untuk itu penulis senantiasa mengharapkan teguran, kritik serta saran yang sifatnya membangun untuk dapat lebih sempurnanya pembuatan tesis atau sejenisnya pada masa-masa yang akan datang. Akhirnya penulis berharap semoga tulisan ini bermanfaat bagi semua pihak, terimakasih.

Palembang, Agustus 2025

Anang Purnomo Kurniawan

DAFTAR ISI

	HALAMAN
COVER DEPAN.....	
COVER DALAM.....	
HALAMAN PENGESAHAN PEMBIMBING	
HALAMAN PENGESAHAN PENGUJI.....	
SURAT PERNYATAAN.....	
ABSTRAK.....	i
<i>ABSTRACT</i>	ii
MOTTO DAN HALAMAN PERSEMBAHAN.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL	vii
DAFTAR GAMBAR	ix
 BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	3
1.3 Batasan Masalah	3
1.4 Rumusan Masalah.....	4
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
1.7 Ruang Lingkup Penelitian.....	6
1.8 Susunan dan Struktur Tesis	6
 BAB II KAJIAN PUSTAKA	
2.1 Kajian Pustaka	8
2.2 Sistem Penerimaan Murid Baru (SPMB).....	12
2.3 Metode K-Nearest Neighbors (KNN).....	19
2.4 Metode Naïve Bayes (NB)	21

BAB III METODOLOGI PENELITIAN

3.1 Waktu Penelitian.....	24
3.2 Alur Penelitian.....	24
3.3 Teknik Pengumpulan Data.....	26
3.4 Sumber Data.....	27
3.5 Alat dan Bahan	28
3.6 Jadwal Penelitian.....	28

BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil Pemetaan SPMB.....	30
4.2 Persiapan data	34
4.3 Perbandingan algoritma KNN dan NB	36
4.4 Penghitungan secara manual	40
4.4.1 Penghitungan K-Nearest Neighbors (KNN).....	40
4.4.2 Penghitungan Naïve Bayes (NB).....	47
4.5 Penghitungan melalui Aplikasi Rapid Minner	56
4.5.1 Alur proses desain dan penghitungan	56
4.5.2 Hasil Penghitungan K-Nearest Neighbors (KNN) kategori pertama.....	60
4.5.3 Hasil Penghitungan Naïve Bayes (NB) kategori pertama.....	61
4.5.4 Hasil Penghitungan K-Nearest Neighbors (KNN) kategori kedua.....	61
4.5.5 asil Penghitungan Naïve Bayes (NB) kategori kedua.....	62

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan.....	62
5.2 Saran.....	62

DAFTAR PUSTAKA

DAFTAR RIWAYAT HIDUP

LAMPIRAN

LEMBAR PERBAIKAN TESIS

DAFTAR TABEL

Tabel 3. 1 Hasil Pendataan Calon Sekolah Sampel.....	25
Tabel 3. 2 Jadwal Penelitian.....	29
Tabel 4. 1 Hasil Pemetaan SPMB per Kabupaten/Kota	31
Tabel 4. 2 Hasil Pemetaan SPMB Kota Palembang.....	32
Tabel 4. 3 Contoh Dataset Kategori Pertama.....	34
Tabel 4. 4 Contoh Dataset Kategori Kedua	35
Tabel 4. 5 Pembagian data	36
Tabel 4. 6 Data Latih Penghitungan Manual KNN	41
Tabel 4. 7 Selisih nilai calon murid dengan data latih.....	44
Tabel 4. 8 Hasil kuadrat data selisih nilai calon murid dengan data latih	44
Tabel 4. 9 Urutan Total Distance.....	45
Tabel 4. 10 Hasil Seleksi Tetangga Terdekat.....	46
Tabel 4. 11 Data Latih Penghitungan Manual Naive Bayes	47
Tabel 4. 12 Hasil Penghitungan Statistik per Mapel	52
Tabel 4. 13 Hasil Penghitungan Posterior Probability per Mapel.....	54

DAFTAR GAMBAR

Gambar 4.1 Penentuan target role sesuai dengan kriteria.....	57
Gambar 4.2 Penentuan split data latih dan data uji	58
Gambar 4.3 Pemilihan algoritma Naïve Bayes	58
Gambar 4.4 Pemilihan algoritma K-Nearest Neighbors	59
Gambar 4.5 Desain penentuan alur penghitungan pada Rapid Minner	60
Gambar 4.6 Hasil Penghitungan KNN versi pertama.....	60
Gambar 4.7 Hasil Penghitungan NB versi pertama.....	61
Gambar 4.8 Hasil Penghitungan KNN versi kedua.....	61
Gambar 4.9 Hasil Penghitungan NB versi kedua.....	61