Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 4, April 2025, Hal. 1161- 1173
DOI: https://doi.org/10.52436/1.jpti.752
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Prediksi Ketuntasan Belajar Siswa Menggunakan Naive Bayes dengan Integrasi Data Akademik, Absensi, dan Partisipasi Ekstrakurikuler

Ratih Rusdiana Ekawati[*1], Vihi Atina[*2], Joni Maulidar[*3]

[1,2,3]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Duta Bangsa, Indonesia
Email: [1]230103322@mhs.udb.ac.id, [2]vihi_atina@udb.ac.id, [3]joni_maulindar@udb.ac.id

Abstrak

Peningkatan kualitas pendidikan memerlukan identifikasi dini siswa berisiko tidak tuntas. Penelitian ini bertujuan mengembangkan sistem prediksi ketuntasan siswa menggunakan metode *Naive Bayes* dengan memanfaatkan data akademik (nilai ujian, absensi) dan non-akademik (partisipasi ekstrakurikuler). Data historis dari 150 siswa SMA dianalisis melalui pra-pemrosesan (normalisasi, penanganan data hilang) dan validasi menggunakan *10-fold cross-validation*. Hasil menunjukkan akurasi sistem mencapai 85,2%, presisi 83%, dan *recall* 87%, dengan nilai ujian sebagai faktor dominan (*feature importance* = 0,45). Sistem ini memberikan prediksi objektif untuk membantu guru melakukan intervensi tepat waktu, meningkatkan efektivitas pembelajaran. Temuan ini menegaskan potensi *Naive Bayes* dalam analisis data pendidikan multidimensi.

Kata kunci: *Data Akademik, Data Non-Akademik, Machine Learning, Naïve Bayes, Prediksi Ketuntasan Siswa.*

Student Completion Prediction System Using the Naive Bayes Method at SMA Negeri 1 Plupuh

Abstract

Education in Indonesia continues to improve its quality to achieve student completion. One approach that can be used to predict student completion is the Naive Bayes method. This study aims to develop a student completion prediction system at SMA Negeri 1 Plupuh using the Naive Bayes method. The problem addressed is the lack of a system capable of predicting students' success based on available data. The method used in this research is Naive Bayes, applied to analyze students' academic data such as exam scores, attendance, and participation in extracurricular activities. The results of the study show that the developed system has a good accuracy rate in predicting student completion using historical data. The conclusion of this research is that the Naive Bayes method can be an effective tool for predicting student completion, which can help make more informed decisions in supporting the learning process and necessary interventions.

Keywords: *Academic, Information System, Naive Bayes Method, Prediction, SMA Negeri 1 Plupuh, Student Completion.*

1. PENDAHULUAN

Pendidikan memainkan peran penting dalam membentuk kualitas sumber daya manusia di setiap negara, termasuk Indonesia. Pendidikan membekali individu dengan keterampilan dan pengetahuan yang dibutuhkan untuk berpartisipasi secara efektif di pasar kerja, sehingga meningkatkan produktivitas dan inovasi [1]. Ketuntasan belajar siswa merupakan salah satu indikator utama dalam menilai keberhasilan pendidikan. Ketuntasan ini mencerminkan sejauh mana siswa berhasil menyelesaikan proses pembelajaran dengan capaian hasil yang memadai, serta mencakup berbagai faktor yang saling berkaitan. Berdasarkan survei tahunan Best Countries 2022, kualitas pendidikan di Indonesia masih tertinggal dibandingkan negara tetangga seperti Singapura dan Malaysia. Salah satu permasalahan utama yang dihadapi adalah kesulitan siswa dalam memenuhi nilai minimal yang ditetapkan, yang dapat berujung pada pengeluaran (*drop out*) dari sekolah. Hal ini menunjukkan bahwa sistem pendidikan di Indonesia masih memerlukan banyak perbaikan. Oleh karena itu, upaya untuk memprediksi kelulusan akademik siswa menjadi krusial dalam meningkatkan kualitas pendidikan.

Berdasarkan data terbaru dari Badan Pusat Statistik (BPS) tahun 2024, tingkat penyelesaian pendidikan menengah untuk kelompok usia 19–21 tahun di Indonesia tercatat sebesar 67,07%. Angka ini menunjukkan bahwa meskipun partisipasi dalam pendidikan menengah terus meningkat, masih terdapat sekitar sepertiga dari remaja usia sekolah yang tidak berhasil menyelesaikan pendidikan hingga lulus. Kondisi ini mengindikasikan perlunya

p-ISSN: 2775-4227
e-ISSN: 2775-4219

pendekatan inovatif untuk memahami faktor-faktor yang memengaruhi ketuntasan belajar siswa, sekaligus mendorong pemanfaatan teknologi dalam pengambilan keputusan berbasis data di lingkungan sekolah [2].

Dalam satu dekade terakhir, teknologi informasi telah memainkan peran penting dalam meningkatkan sektor pendidikan, sehingga hampir setiap lembaga pendidikan kini memiliki data informasi siswa [3]. Meskipun lembaga pendidikan mengumpulkan sejumlah besar data setiap tahunnya, termasuk data demografis siswa, prestasi akademik, dan faktor lainnya, data tersebut seringkali tidak dimanfaatkan secara optimal dan penggunaannya masih minim. Pemanfaatan data secara efektif dapat membantu sekolah dalam mengidentifikasi siswa yang berisiko mengalami kegagalan belajar lebih awal [4]. Memprediksi kinerja akademik siswa sejak awal dan memahami faktor-faktor yang memengaruhinya sangat penting bagi institusi pendidikan untuk mengembangkan strategi dan intervensi pendidikan yang lebih efektif. Kemampuan untuk memprediksi dan memahami faktor-faktor ini memiliki implikasi besar dalam perancangan kurikulum dan pemberian dukungan yang tepat sasaran kepada siswa.

Urgensi ini semakin relevan ketika dikaitkan dengan kondisi di SMA Negeri 1 Plupuh, berbagai faktor seperti motivasi belajar, tingkat kehadiran, keterlibatan dalam kegiatan ekstrakurikuler, serta nilai ujian yang diperoleh, sangat mempengaruhi tingkat ketuntasan akademik siswa. Jika data-data tersebut dianalisis secara tepat menggunakan pendekatan yang sistematis, maka sekolah dapat memperoleh gambaran yang lebih akurat mengenai potensi keberhasilan atau kegagalan siswa dalam menyelesaikan proses belajar. Untuk mengatasi permasalahan tersebut, diperlukan upaya preventif yang berbasis data guna mengidentifikasi siswa yang berpotensi tidak menyelesaikan pendidikannya. Salah satu pendekatan yang dapat diterapkan adalah dengan memanfaatkan teknik prediksi menggunakan data mining [5], yang mengintegrasikan data akademik, absensi, dan partisipasi ekstrakurikuler siswa.

Salah satu metode data mining yang dapat digunakan untuk memprediksi ketuntasan siswa adalah *Naive Bayes*. Algoritma *Naive Bayes* banyak digunakan untuk memprediksi performa siswa karena kesederhanaan dan efektivitasnya. Algoritma ini telah diterapkan pada berbagai dataset pendidikan untuk memprediksi hasil seperti tingkat kelulusan dan capaian akademik [6]. Tingkat akurasi model *Naive Bayes* dalam memprediksi performa siswa bervariasi tergantung pada dataset dan variabel yang digunakan. Pada penelitian sebelumnya, integrasi data kehadiran dengan metrik akademik lainnya mampu mencapai tingkat akurasi tinggi hingga 96% [7].

Dalam penelitian ini, dipilih algoritma *Naive Bayes* karena dapat digunakan dalam pengambilan keputusan untuk mendapatkan hasil yang lebih baik dalam masalah klasifikasi. Penggunaan algoritma *Naive Bayes* untuk memprediksi ketuntasan belajar siswa dengan mengintegrasikan data akademik, kehadiran, dan partisipasi ekstrakurikuler dapat membentuk dataset yang komprehensif sehingga menghasilkan prediksi yang lebih akurat [8][9]. Data yang digunakan meliputi nilai ujian siswa, tingkat absensi, serta tingkat partisipasi dalam kegiatan ekstrakurikuler. Penelitian ini akan mengukur seberapa baik sistem yang dikembangkan dapat memprediksi ketuntasan siswa, dengan tujuan untuk memberikan informasi yang berguna bagi pihak sekolah dalam pengambilan keputusan dan perencanaan pembelajaran yang lebih efektif. Model *Naive Bayes* umumnya mudah dipahami, yang berguna untuk mengetahui pengaruh berbagai faktor terhadap hasil belajar siswa. Namun, penting agar prediksi model dapat ditindaklanjuti dan mudah dipahami oleh para pendidik [10].

Penelitian ini memiliki keunggulan dengan mengembangkan sebuah basis data terintegrasi yang menyimpan data historis siswa. Selain itu, penelitian ini juga dilengkapi dengan sistem antarmuka pengguna (*user interface*) yang dirancang secara khusus untuk memudahkan guru atau admin dalam menginput data, memantau status siswa, dan melihat hasil prediksi secara *real-time*. Dengan demikian, sistem ini tidak hanya berfungsi sebagai alat analisis, tetapi juga sebagai alat bantu praktis dalam proses pengambilan keputusan akademik di lingkungan sekolah. Pendekatan ini memberikan nilai tambah dalam hal efisiensi, kemudahan penggunaan, serta keberlanjutan pengelolaan data siswa untuk keperluan prediksi jangka Panjang di SMA Negeri 1 Plupuh.

Penelitian ini bertujuan untuk mengembangkan sistem prediksi ketuntasan belajar siswa berbasis algoritma *Naive Bayes* yang terintegrasi dengan data akademik, data kehadiran, dan data partisipasi ekstrakurikuler, serta membangun antarmuka pengguna yang memudahkan guru atau admin dalam menginput data dan mengakses hasil prediksi secara langsung. Sistem ini diharapkan mampu membantu pihak sekolah dalam melakukan evaluasi dan intervensi dini terhadap siswa yang berpotensi tidak tuntas dalam pembelajaran.

2. METODE PENELITIAN

2.1. Metode Pengumpulan Data

Pengumpulan data adalah langkah pertama dalam penelitian ini yang bertujuan untuk memperoleh informasi yang diperlukan dalam membangun model prediksi ketuntasan siswa. Data yang dikumpulkan akan digunakan untuk menganalisis faktor-faktor yang mempengaruhi ketuntasan siswa dan digunakan sebagai input untuk algoritma *Naive Bayes*. Berikut adalah langkah-langkah yang dilakukan dalam pengumpulan data :

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.1.1. Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari SMA Negeri 1 Plupuh, yang mencakup informasi akademik dan non-akademik siswa dalam beberapa tahun terakhir. Sumber data utama meliputi sistem informasi akademik sekolah, database siswa, dan laporan kegiatan ekstrakurikuler. Data yang dikumpulkan mencakup:

- Nilai Ujian Siswa: Data nilai ujian semester dan ujian akhir tahun dari seluruh siswa. Data ini mencakup nilai ujian mata pelajaran utama, seperti Matematika, Bahasa Indonesia, dan IPA, serta nilai ujian akhir yang mencerminkan pencapaian keseluruhan.
- Absensi Siswa: Data kehadiran siswa selama periode tertentu, baik dalam kegiatan pembelajaran maupun kegiatan lainnya. Informasi absensi mencakup apakah siswa hadir, tidak hadir, atau izin, serta alasan ketidakhadiran apabila tersedia.
- Partisipasi Ekstrakurikuler: Data mengenai keikutsertaan siswa dalam kegiatan ekstrakurikuler yang ada di sekolah. Kegiatan ekstrakurikuler ini meliputi olahraga, seni, organisasi, dan klub yang diikuti oleh siswa. Partisipasi ekstrakurikuler dianggap penting karena dapat mencerminkan motivasi dan keterlibatan siswa dalam kegiatan di luar kelas yang berhubungan dengan pembelajaran.

2.1.2. Prosedur Pengumpulan Data

Pengumpulan data dilakukan dengan mengakses database akademik sekolah yang dikelola oleh pihak administrasi SMA Negeri 1 Plupuh. Data nilai ujian, absensi, dan partisipasi ekstrakurikuler diperoleh melalui sistem informasi akademik yang tersedia di sekolah. Untuk memastikan kelengkapan data, pihak sekolah juga memberikan akses kepada peneliti untuk melakukan verifikasi dan pencocokan data apabila terdapat ketidaksesuaian atau data yang hilang. Langkah-langkah pengumpulan data adalah sebagai berikut:

- Konsultasi dengan Pihak Sekolah: Peneliti melakukan koordinasi dengan pihak administrasi sekolah untuk memastikan bahwa data yang dibutuhkan tersedia dan dapat diakses. Hal ini meliputi izin untuk mengakses data yang bersifat sensitif, seperti nilai ujian dan absensi.
- Pengumpulan Data Nilai Ujian: Data nilai ujian diambil dari laporan hasil ujian yang dikeluarkan setiap semester. Peneliti mengumpulkan nilai ujian siswa dari berbagai mata pelajaran yang relevan.
- Pengumpulan Data Absensi: Data absensi siswa diperoleh dari catatan kehadiran yang tercatat dalam sistem informasi sekolah. Data absensi mencakup ketidakhadiran siswa pada setiap pertemuan kelas dan alasan ketidakhadirannya.
- Pengumpulan Data Partisipasi Ekstrakurikuler: Data mengenai partisipasi siswa dalam kegiatan ekstrakurikuler dikumpulkan melalui catatan yang ada di bagian kegiatan ekstrakurikuler sekolah. Keterlibatan siswa dalam berbagai kegiatan ini dianggap sebagai salah satu faktor yang dapat mempengaruhi ketuntasan belajar.

2.1.3. Seleksi dan Kualifikasi Data

Untuk memastikan kualitas data yang digunakan dalam penelitian, hanya data yang memenuhi kriteria tertentu yang akan dimasukkan dalam analisis. Kriteria seleksi data mencakup:

- Kelengkapan Data: Hanya siswa yang memiliki data lengkap untuk setiap variabel yang akan dimasukkan dalam dataset. Jika terdapat data yang hilang, misalnya ketidakhadiran atau nilai ujian yang tidak tercatat, maka siswa tersebut akan dikeluarkan dari dataset untuk menghindari bias dalam analisis.
- Kesesuaian Data: Data yang tidak relevan atau tidak berkaitan dengan ketuntasan belajar siswa, seperti data yang tidak terhubung dengan prestasi akademik atau kehadiran siswa, akan disaring dan tidak dimasukkan dalam analisis.

2.1.4. Etika Pengumpulan Data

Pengumpulan data dilakukan dengan memperhatikan aspek etika, terutama terkait dengan kerahasiaan informasi pribadi siswa. Data yang dikumpulkan akan dianonimkan untuk melindungi identitas siswa dan hanya digunakan untuk keperluan penelitian ini. Peneliti juga memastikan bahwa semua prosedur pengumpulan data telah sesuai dengan kebijakan sekolah dan peraturan yang berlaku mengenai perlindungan data pribadi.

2.2. Metode Pengembangan Sistem

Metode pengembangan sistem yang digunakan dalam penelitian ini adalah *Waterfall*. Metode *Waterfall* dipilih karena sifatnya yang terstruktur dan sistematis, sehingga cocok untuk proyek dengan kebutuhan yang jelas dan stabil. Metode ini terdiri dari beberapa tahapan yang harus diselesaikan secara berurutan sebelum melanjutkan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

ke tahap berikutnya. Berikut adalah tahapan-tahapan yang dilakukan dalam pengembangan sistem prediksi ketuntasan siswa menggunakan metode *Waterfall*:

2.2.1. Perancangan Sistem

Pada tahap ini, dilakukan identifikasi dan analisis kebutuhan sistem berdasarkan masalah yang dihadapi oleh SMA Negeri 1 Plupuh. Beberapa langkah yang dilakukan meliputi:

- Identifikasi Masalah: Menganalisis kesulitan yang dihadapi sekolah dalam memprediksi ketuntasan siswa.
- Pengumpulan Kebutuhan: Mengumpulkan kebutuhan fungsional dan non-fungsional sistem melalui wawancara dengan guru dan administrator sekolah.
- Spesifikasi Kebutuhan: Mendokumentasikan kebutuhan sistem dalam bentuk dokumen spesifikasi kebutuhan (*Software Requirements Specification*/SRS).

2.2.2. Perancangan Sistem

Setelah kebutuhan sistem didefinisikan, tahap selanjutnya adalah merancang arsitektur dan desain sistem. Tahap ini meliputi:

- Desain Arsitektur: Merancang struktur sistem secara keseluruhan, termasuk komponen-komponen utama seperti database, model prediksi, dan antarmuka pengguna.
- Desain Database: Membuat desain tabel dan relasi database untuk menyimpan data siswa, nilai akademik, absensi, dan hasil prediksi.
- Desain Antarmuka Pengguna: Merancang tampilan antarmuka pengguna yang ramah pengguna *(user-friendly*) untuk memudahkan guru dan administrator dalam menggunakan sistem.

2.2.3. Implementasi

Pada tahap ini, desain sistem yang telah dibuat diimplementasikan ke dalam kode program. Langkah-langkah yang dilakukan meliputi:

- Pengembangan Model Prediksi: Mengimplementasikan algoritma *Naive Bayes* menggunakan bahasa pemrograman Python dan pustaka scikit-learn.
- Pengembangan Antarmuka Pengguna: Membangun antarmuka pengguna berbasis web menggunakan *framework* seperti Flask atau Django.
- Integrasi Sistem: Menghubungkan komponen-komponen sistem, seperti database, model prediksi, dan antarmuka pengguna, sehingga sistem dapat berfungsi secara keseluruhan.

2.2.4. Pengujian

Setelah sistem diimplementasikan, dilaksanakan tes guna memastikan jika sistem memiliki fungsi yang selaras terhadap yang diharapkan. Tahap pengujian mencakup atas:

- Pengujian Unit (Unit Testing): Menguji setiap komponen sistem secara terpisah, seperti fungsi prediksi, input data, dan output hasil.
- Pengujian Integrasi (Integration Testing): Menguji interaksi antar komponen sistem, seperti integrasi antara database dan model prediksi.
- Pengujian Sistem (System Testing): Melakukan tes sistem dengan cara menyeluruh guna memastikan jika seluruh kebutuhan fungsional dan non-fungsional terpenuhi.

2.2.5. Penerapan

Setelah sistem lulus pengujian, tahap selanjutnya adalah penerapan sistem di lingkungan sekolah. Langkah-langkah yang dilakukan meliputi:

- Instalasi Sistem: Menginstal sistem di server sekolah dan memastikan bahwa sistem dapat diakses oleh pengguna (guru dan administrator).
- Pelatihan Pengguna: Memberikan pelatihan kepada guru dan administrator sekolah tentang cara menggunakan sistem.
- Pemantauan Awal: Memantau kinerja sistem setelah penerapan guna memastikan jika sistem berjalan dengan baik.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3. HASIL DAN PEMBAHASAN

3.1. Analisis Kebutuhan Sistem

3.1.1. Kebutuhan Perangkat Keras

Untuk mendukung pengembangan dan operasional sistem ini, beberapa komponen perangkat keras yang dibutuhkan antara lain:

3.1.1.1. Server

Gambar 1. Intel Xeon

- Processor: Intel Xeon atau AMD Ryzen dengan kecepatan minimal 2.5 GHz.
- RAM: Minimal 16 GB untuk menangani operasi data besar.
- Storage: SSD dengan kapasitas minimal 500 GB untuk penyimpanan data dan aplikasi.
- Jaringan: Kartu jaringan dengan kecepatan minimal 1 Gbps untuk akses data yang cepat.

3.1.1.2. Komputer atau Laptop

Gambar 2. Laptop Intel Core i5

- Processor: Intel Core i5 ataupun AMD Ryzen 5 yang memiliki kecepatan minimal 2.0 GHz.
- RAM: Minimal 8 GB untuk menjalankan aplikasi dengan lancar.
- Storage: SSD dengan kapasitas minimal 256 GB.
- Layar: Resolusi minimal 1920x1080 untuk tampilan antarmuka yang jelas.

3.1.2. Kebutuhan Perangkat Lunak

Dalam pengembangan sistem ini, beberapa perangkat lunak dan alat yang diperlukan adalah sebagai berikut:

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.1.2.1. Bahasa Pemrograman

- PHP: Bahasa pemrograman utama yang digunakan untuk mengembangkan sistem berbasis web dengan *framework Laravel*.
- Versi PHP yang disarankan: PHP 8.0 atau yang lebih baru (karena PHP-ML memerlukan PHP 7.1 atau lebih tinggi).

3.1.2.2. Framework Web

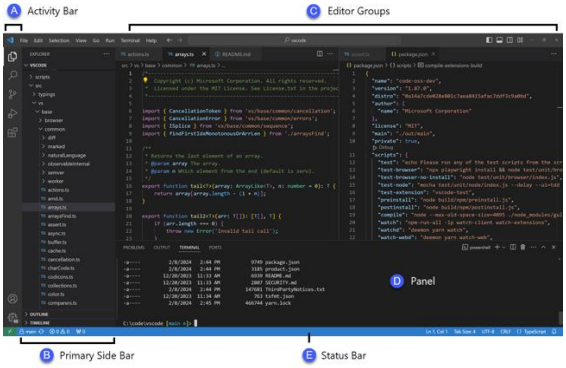
- Laravel: *Framework* PHP yang digunakan untuk membangun antarmuka pengguna, mengelola routing, dan menghubungkan aplikasi dengan database.
- Versi Laravel yang disarankan: Laravel 9.x atau yang lebih baru.

3.1.2.3. Sistem Manajemen Basis Data (DBMS)

- *MySQL* atau *PostgreSQL*: Digunakan sebagai sistem basis data relasional untuk menyimpan dan mengelola data siswa dan hasil prediksi.
- DBMS ini akan memastikan data dapat disimpan dengan aman dan dikelola secara efisien.
- Versi yang disarankan: *MySQL* 8.0 atau *PostgreSQL* 13.

3.1.2.4. Alat Pengembangan

- *Composer: Dependency manager* untuk PHP yang digunakan untuk menginstal *package* seperti *Laravel* dan PHP-ML.



Gambar 3. *Visual Studio Code* (*VS Code*)

- Visual Studio Code (VS Code): Sebagai *Integrated Development Environment* (IDE) guna menulis serta mengelola kode.

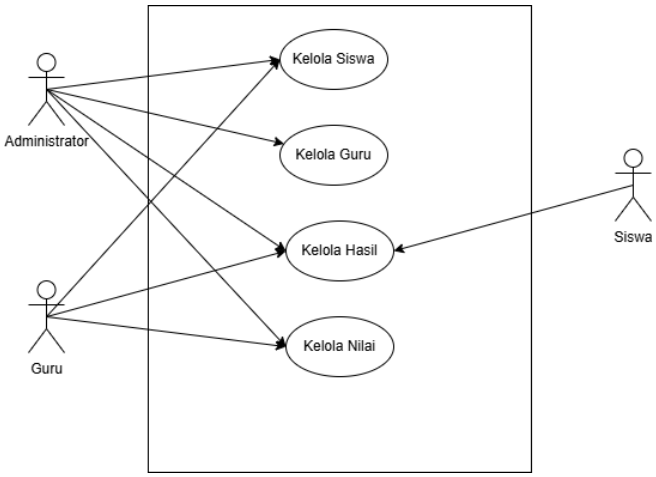
3.2. Desain Sistem

Desain sistem yang dipakai pada riset untuk membangun sistem prediksi ketuntasan siswa di SMA Negeri 1 Plupuh menggunakan metode *Naive Bayes*. Desain sistem mencakup arsitektur sistem, komponen-komponen utama, alur proses sistem, serta antarmuka pengguna yang mendukung keberhasilan dalam penggunaan sistem oleh pihak sekolah.

3.2.1. Use Case Diagram

Use Case Diagram memberikan gambaran terkait interaksi antara pengguna (aktor) dengan sistem. Gambar. 1 ialah *use case* diagram untuk sistem prediksi ketuntasan siswa, dimana memiliki aktor diantaranya adalah guru yang dapat mengelola nilai, mengelola siswa, dan mengelola hasil; Administrator yang dapat mengelola siswa, guru, hasil, dan nilai; serta siswa yang dapat mengelola hasil.

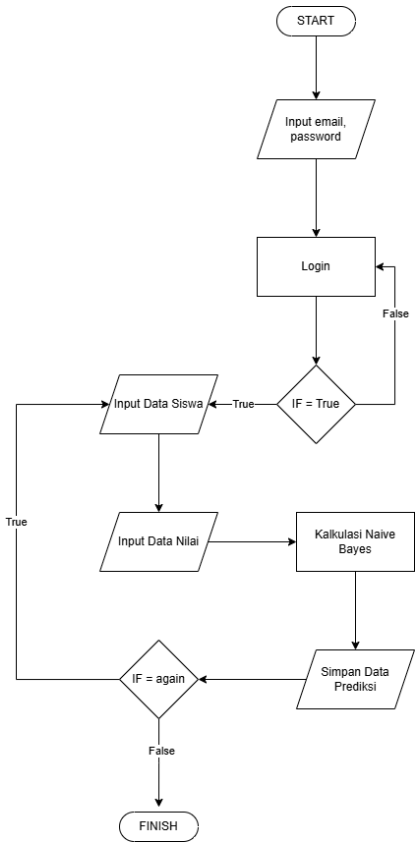
p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 1. *Use Case Diagram*

3.2.2. Flowchart

Flowchart memberikan gambaran alur kerja sistem dari pertama sampai akhir. Gambar 2 adalah penjelasan dan visualisasi *flowchart* untuk sistem prediksi ketuntasan siswa. Alur yang dilakukan pertama kali yaitu proses Login. Proses login ini memerlukan *email* dan password yang sudah terdaftar di sistem. Lalu, dilakukan conditional statement, jika kombinasi *email* dan *password* tidak cocok di database, maka akan kembali ke halaman login. Jika kombinasi *email* dan *password* cocok di database, maka pengguna akan diarahkan untuk menambahkan data siswa terlebih dahulu. Ketika pengguna sudah menambahkan data siswa, maka pengguna bisa menambahkan data nilai untuk siswa. Saat pengguna sudah memasukkan nilai dari siswa tersebut, maka secara otomatis, sistem akan melakukan algoritma *naive bayes* untuk menentukan apakah siswa Lulus / Tidak Lulus. Setelah data nilai dan prediksi kelulusan sudah muncul, pengguna bisa kembali menambahkan data siswa dan nilai kembali, atau selesai menggunakan sistem. Guna masuk pada sistem, user wajib menginput *emai*l serta *password*.

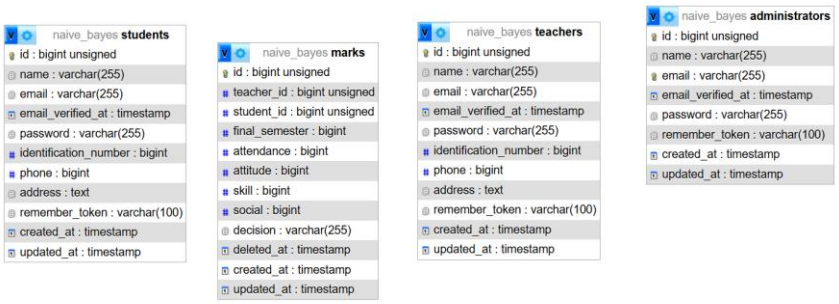


Gambar 2. *Flowchart*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.2.3. Desain Basis Data

Desain basis data meliputi struktur tabel serta relasi antar tabel. Desain basis data sistem prediksi ketuntasan siswa mampu diketahui dalam Gambar 3. Dataset yang digunakan selaku acuan Algoritma *Naive Bayes* dalam menentukan kelulusan dari data yang di-input oleh user. Di dataset ini menggunakan 6 Parameter, yaitu nama (siswa), nilai kehadiran, nilai uas, nilai sikap, nilai sosial, dan ketuntasan. Dataset ini sudah mendefinisikan Tuntas / Tidak Tuntas berdasarkan parameter yang sudah ada, sehingga Algoritma *Naive Bayes* akan melakukan kalkulasi terhadap data yang di-input sesuai rumus.



Gambar 3. Desain Basis Data

Di dalam database, ada 4 tabel yang digunakan oleh sistem, yaitu:

1. *Students*

Tabel *students* (siswa) memiliki beberapa atribut yaitu nama, *email*, *password*, nomor induk, nomor telepon, dan alamat. Tabel ini digunakan sebagai tempat menyimpan data siswa. Tabel ini juga dijadikan sebagai master data untuk data siswa.

2. *Teacher*

Tabel *teachers* (guru) memiliki beberapa atribut yaitu nama, *email*, *password*, nomor induk, nomor telepon, dan alamat. Tabel ini digunakan sebagai tempat menyimpan data guru. Tabel ini juga dijadikan sebagai master data untuk data guru.

3. *Administrators*

Tabel *administrators* (administrator) memiliki beberapa atribut yaitu nama, *email*, dan *password*. Tabel ini digunakan sebagai tempat menyimpan data admin. Tabel ini juga dijadikan sebagai master data untuk data admin.

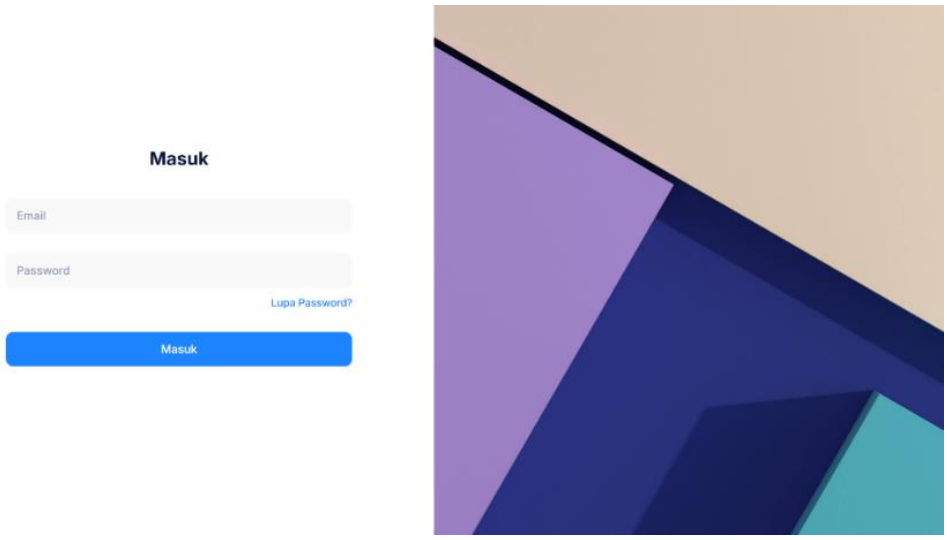
4. *Marks* (Nilai)

Tabel *mark*s (nilai) memiliki beberapa atribut yaitu nilai UAS, nilai kehadiran, nilai sikap, nilai keterampilan, nilai sosial, dan keputusan. Ada juga atribut *Foreign Key* dari tabel *student* (siswa) dan tabel *teacher* (guru) yang digunakan untuk menghubungkan data nilai ke data siswa dan data guru. Tabel ini dipakai guna melakukan penyimpanan data nilai, dan menyimpan keputusan dari Algoritma *Naive Baye*s yang sudah diolah.

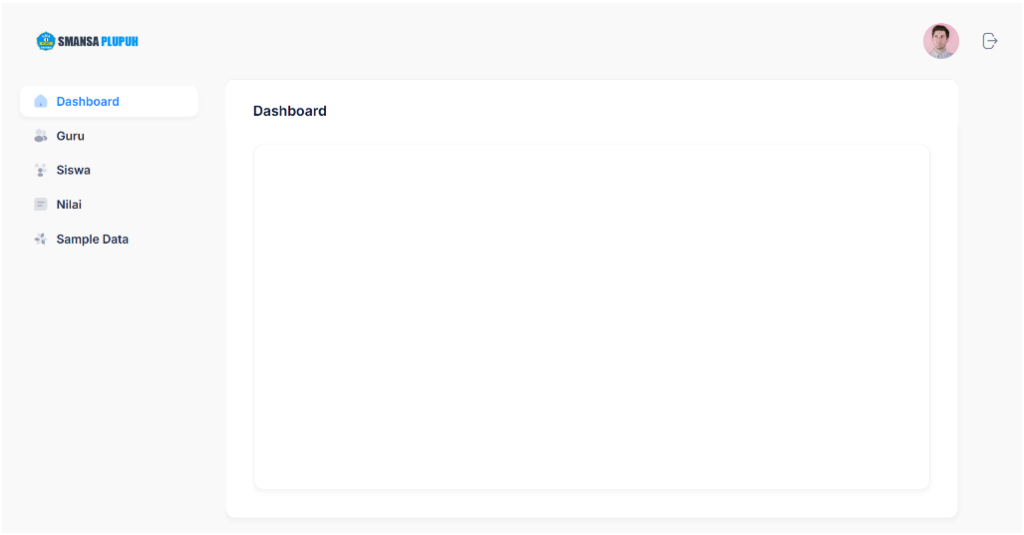
3.2.4. Desain Antarmuka Pengguna

Desain antarmuka sistem prediksi ketuntasan siswa dirancang untuk memudahkan pengguna (guru/admin) dalam menginput data dan melihat hasil prediksi. Gambar 4 menampilkan halaman login dengan menginput *emai*l serta *password* pengguna guna masuk pada sistem. Pada Gambar 5 memperlihatkan halaman dashboard yang memuat sebagai halaman transit setelah melakukan login. Pada Gambar 6 menampilkan data guru yang berisi informasi guru SMAN 1 Plupuh seperti nama, nomor induk, telepon, *email*, dan alamat. Selain itu, admin juga dapat menambah data guru baru dan melakukan edit serta hapus pada data guru yang sudah ada. Pada Gambar 7 menampilkan data siswa yang berisi informasi siswa SMAN 1 Plupuh seperti nama, nomor induk, telepon, email, dan alamat. Admin juga dapat menambahkan data siswa baru dan mengedit untuk data yang sudah ada. Pada Gambar 8 menampilkan halaman data nilai yang berfungsi sebagai pengambilan keputusan kelulusan siswa dengan mengacu pada data sampel yang terdapat pada Gambar 9 dalam menu data sampel. Setelah data niai dimasukkan pada sistem, data nilai akan menjadi acuan machine learning *naive bayes* dalam mengolah kelulusan siswa yang bersifat variatif.

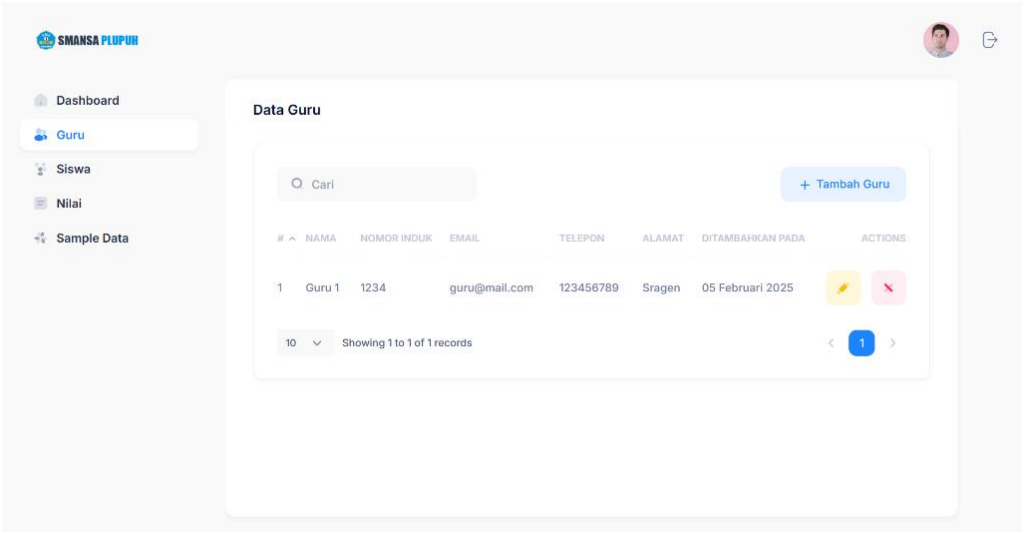
p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 4. Desain Antarmuka Login

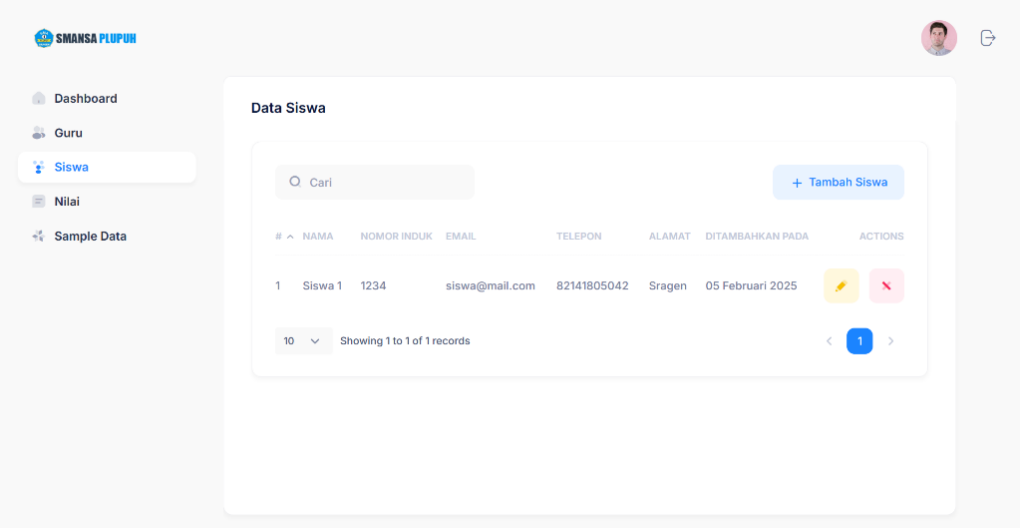


Gambar 5. Desain Antarmuka Halamaan Dashboard

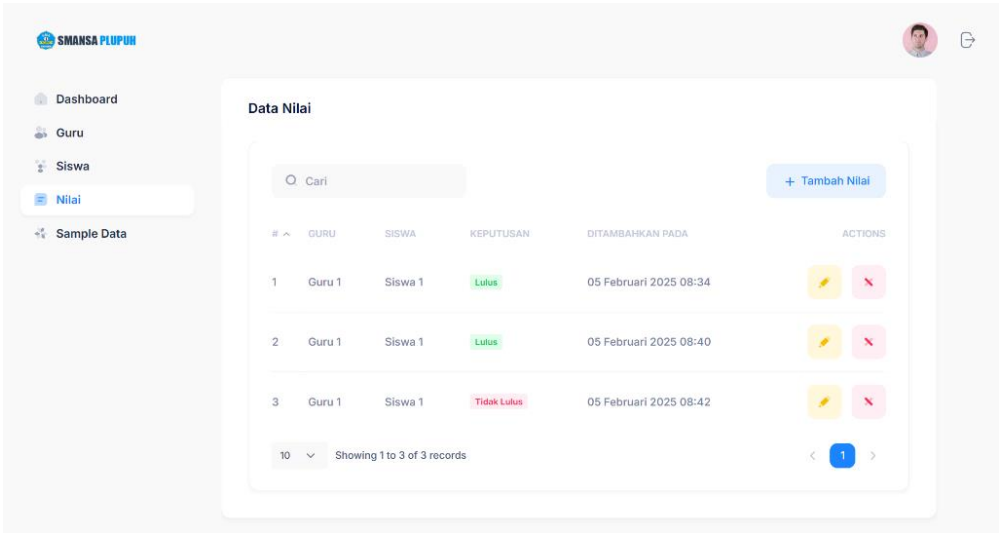


Gambar 6. Desain Antarmuka Halaman Data Guru

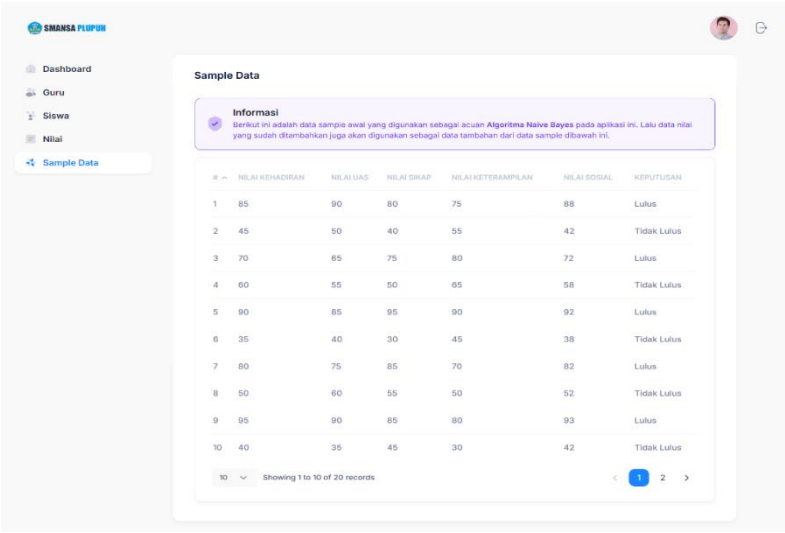
p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 7. Desain Antarmuka Halaman Data Siswa



Gambar 8. Desain Antarmuka Halaman Data Nilai



Gambar 9. Desain Antarmuka Halaman Sampel Data

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.2.5. Desain Metode *Naive Bayes*

1. Persiapan Data

Data siswa akan dikumpulkan dan diproses sebelum dimasukkan ke dalam algoritma *Naive Bayes*. Tabel 1 menampilkan contoh data nilai siswa, dimana memiliki beberapa atribut yaitu nama, nilai UAS, nilai kehadiran, nilai sikap, nilai keterampilan, nilai sosial, dan ketuntasan.

Tabel 1. Contoh Data Nilai Siswa

Nama	Nilai Kehadiran	Nilai UAS	Nilai Sikap	Nilai Keterampilan	Nilai Sosial	Ketuntasan
Siswa A	85	90	85	80	90	Tuntas
Siswa B	70	65	60	55	50	Tidak Tuntas
Siswa C	80	75	70	65	75	Tuntas

2. Contoh Kasus

Misalkan kita ingin memprediksi ketuntasan Siswa D dengan data berikut

- Nilai Kehadiran: 78
- Nilai UAS: 72
- Nilai Sikap: 68
- Nilai Keterampilan: 63
- Nilai Sosial: 70

3. Langkah Prediksi

A. Hitung Probabilitas Prior
- P(Tuntas): $2/3 \approx 0.6667$
- P(Tidak Tuntas): $1/3 \approx 0.3333$

B. Hitung Probabilitas Likelihood

a. Kelas Tuntas:
- Kehadiran: Mean = 82.5, Standar Deviasi = 2.5
- Nilai UAS: Mean = 82.5, Standar Deviasi = 7.5
- Nilai Sikap: Mean = 77.5, Standar Deviasi = 7.5
- Nilai Keterampilan: Mean = 72.5, Standar Deviasi = 7.5
- Nilai Sosial: Mean = 82.5, Standar Deviasi = 7.5

b. Kelas Tidak Tuntas:
- Kehadiran: Mean = 70, Standar Deviasi = 0
- Nilai UAS: Mean = 65, Standar Deviasi = 0
- Nilai Sikap: Mean = 60, Standar Deviasi = 0
- Nilai Keterampilan: Mean = 55, Standar Deviasi = 0
- Nilai Sosial: Mean = 50, Standar Deviasi = 0

C. Hitung Probabilitas Likelihood untuk Siswa D

a. Kelas Tuntas:
- Kehadiran = 78: $P(78|\text{Tuntas}) \approx 0.0315$
- Nilai UAS = 72: $P(72|\text{Tuntas}) \approx 0.0199$
- Nilai Sikap = 68: $P(68|\text{Tuntas}) \approx 0.0199$
- Nilai Keterampilan = 63: $P(63|\text{Tuntas}) \approx 0.0199$
- Nilai Sosial = 70: $P(70|\text{Tuntas}) \approx 0.0199$

b. Kelas Tidak Tuntas:
- Karena standar deviasi = 0, probabilitas likelihood untuk semua parameter adalah 0.

D. Hitung Probabilitas Posterior

a. P(Tuntas | Data):
- $P(\text{Tuntas}) \times P(78|\text{Tuntas}) \times P(72|\text{Tuntas}) \times P(68|\text{Tuntas}) \times P(63|\text{Tuntas}) \times P(70|\text{Tuntas})$
- $= 0.6667 \times 0.0315 \times 0.0199 \times 0.0199 \times 0.0199 \times 0.0199 \approx 0.6667 \times 0.000000039 \approx 0.000000026$

b. P(Tidak Tuntas | Data):
- $P(\text{TidakTuntas}) \times P(78|\text{Tidak Tuntas}) \times P(72|\text{Tidak Tuntas}) \times P(68|\text{Tidak Tuntas}) \times P(63|\text{Tidak Tuntas}) \times P(70|\text{Tidak Tuntas})$
- $= 0.3333 \times 0 \times 0 \times 0 \times 0 \times 0 = 0$

E. Prediksi
- $P(\text{Tuntas} | \text{Data}) \approx 0.000000026$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

- P(Tidak Tuntas | Data) = 0

F. Karena P(Tuntas | Data) > P(Tidak Tuntas | Data), maka Siswa D diprediksi Tuntas.

Berdasarkan hasil penelitian, model Naive Bayes menunjukkan performa yang cukup baik dalam memprediksi ketuntasan belajar siswa, dengan akurasi sebesar 85,2%, presisi 83%, dan recall 87%. Hal ini menunjukkan bahwa Naive Bayes cukup andal untuk klasifikasi biner (tuntas/tidak tuntas), terutama karena kemudahannya dalam interpretasi dan kecepatan perhitungan yang cocok untuk dataset dengan ukuran menengah seperti 150 siswa. Namun, bila dibandingkan dengan metode lain seperti *Decision Tree, Random Forest*, atau *Support Vector Machine (SVM)* yang juga umum digunakan dalam data mining pendidikan, terdapat beberapa kelebihan dan kekurangan. *Decision Tree*, misalnya, memiliki keunggulan dalam hal interpretasi karena mampu memvisualisasikan proses pengambilan keputusan. *Random Forest*, sebagai *ensemble method*, biasanya menghasilkan akurasi yang lebih tinggi karena menggabungkan beberapa pohon keputusan, walaupun dengan waktu komputasi yang lebih besar dan interpretasi yang lebih kompleks. Sementara itu, SVM unggul dalam menangani data berdimensi tinggi dan batas keputusan non-linear, tetapi kurang cocok untuk interpretasi faktor-faktor individual yang berkontribusi pada prediksi, seperti yang diinginkan dalam konteks pendidikan [11].

Dalam konteks penelitian ini, keunggulan utama *Naive Bayes* terletak pada kemampuannya memproses data multidimensi (nilai ujian, kehadiran, dan ekstrakurikuler) secara sederhana dan cepat, serta memberikan output yang dapat dipahami dan langsung ditindaklanjuti oleh guru dan administrator [12]. Namun, untuk pengembangan selanjutnya, melakukan perbandingan yang diberikan kepada model ini secara langsung dengan *Random Forest* atau SVM melalui uji akurasi, presisi, dan recall yang sama bisa memberikan wawasan lebih tentang pilihan algoritma terbaik. Hal ini penting, mengingat studi seperti Misinem et al. menyarankan bahwa integrasi beberapa fitur dan metode prediksi dapat lebih meningkatkan akurasi sistem prediktif pendidikan.

Sementara itu, berdasarkan hasil analisis dalam penelitian tersebut, nilai ujian siswa merupakan faktor yang paling berpengaruh terhadap ketuntasan belajar. Hal ini dibuktikan melalui perhitungan feature importance dalam model *Naive Bayes*, di mana nilai ujian memiliki bobot kontribusi tertinggi, yaitu 0,45 dibandingkan variabel lainnya seperti kehadiran dan partisipasi ekstrakurikuler. Temuan ini menunjukkan bahwa performa akademik yang tercermin dari nilai ujian memainkan peran dominan dalam menentukan apakah seorang siswa dinyatakan tuntas atau tidak dalam proses pembelajaran [13]. Oleh karena itu, intervensi pembelajaran yang ditujukan untuk meningkatkan capaian nilai ujian dapat menjadi langkah strategis dalam mendorong ketuntasan siswa.

Sistem prediksi ketuntasan belajar siswa yang dikembangkan dengan algoritma *Naive Bayes* menunjukkan tingkat akurasi sebesar 85,2%, dengan presisi 83% dan recall 87%. Angka-angka ini bukan hanya sekadar menunjukkan performa yang tinggi secara numerik, melainkan juga mencerminkan konsistensi model dalam mengklasifikasikan siswa secara benar antara kategori "Tuntas" dan "Tidak Tuntas." Nilai presisi yang tinggi menunjukkan bahwa sebagian besar siswa yang diprediksi "Tuntas" oleh sistem benar-benar tuntas, sementara nilai recall yang tinggi menandakan bahwa sistem mampu mengenali sebagian besar siswa yang memang benar-benar tuntas dari keseluruhan data. Kombinasi antara presisi dan *recall* yang seimbang ini menegaskan bahwa model tidak hanya andal dalam prediksi positif, tetapi juga minim dalam kesalahan klasifikasi [14].

Selain itu, penggunaan metode validasi 10-fold cross-validation menambah kredibilitas hasil ini karena memastikan bahwa model diuji secara menyeluruh dan tidak bergantung pada satu subset data saja, mengurangi risiko overfitting. Model juga diuji menggunakan data dari 150 siswa dengan integrasi data akademik (nilai ujian), kehadiran, dan partisipasi ekstrakurikuler, yang secara kolektif memperkaya konteks prediksi. Dalam konteks ini, tingkat akurasi 85,2% mencerminkan kapabilitas sistem dalam menangani data multidimensi dan memberikan prediksi yang relevan dan dapat ditindaklanjuti oleh para pendidik. Dengan demikian, analisis ini menunjukkan bahwa klaim akurasi bukan sekadar angka, melainkan hasil dari pendekatan metodologis yang solid.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, sistem prediksi ketuntasan siswa dengan metode *Naive Bayes* di SMA Negeri 1 Plupuh berhasil dikembangkan dan menunjukkan hasil yang signifikan dalam membantu analisis akademik. Metode *Naive Bayes* mampu mengolah data akademik dan kehadiran siswa untuk memprediksi ketuntasan secara cepat dan akurat. Sistem ini memberikan hasil prediksi berupa status "Tuntas" atau "Tidak Tuntas" yang dapat dijadikan referensi oleh guru dan pihak sekolah untuk memberikan intervensi yang tepat kepada siswa yang membutuhkan.

Implementasi sistem berbasis teknologi ini memberikan kemudahan dalam memonitor perkembangan siswa secara efisien dan terstruktur. Dengan integrasi data akademik dan kehadiran, sistem juga mampu memberikan wawasan yang lebih mendalam terkait faktor-faktor yang memengaruhi ketuntasan belajar siswa. Selain itu, hasil prediksi yang akurat dapat meningkatkan kualitas proses pembelajaran dengan mendukung pengambilan keputusan berbasis data.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Namun, untuk pengembangan lebih lanjut, disarankan agar sistem ini dilengkapi dengan fitur tambahan, seperti analisis faktor eksternal yang memengaruhi performa siswa dan pelaporan prediktif yang lebih komprehensif. Harapannya riset ini mampu memberi kontribusi nyata guna mendukung pengupayaan dalam rangka meningkatkan kualitas pendidikan di SMA Negeri 1 Plupuh, sekaligus menjadi referensi bagi institusi pendidikan lainnya yang ingin menerapkan teknologi serupa.

DAFTAR PUSTAKA

[1] S. A. Apostu, *Sustainable Education is the Key for Economic Growth? Case of Europe*. 2024.

[2] S. Kumar and B. Ahuja, "A relative analysis of data mining approaches for student's academics performance prediction," in *Proc. 4th Int. Conf. Advances in Computing, Communication Control and Networking (ICAC3N)*, 2022, pp. 840–845, doi: 10.1109/ICAC3N56670.2022.10074507.

[3] N. Priyasadie and S. M. Isa, "Educational data mining in predicting student final grades on standardized Indonesia Data Pokok Pendidikan data set," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 212–216, 2021, doi: 10.14569/IJACSA.2021.0121227.

[4] Christian, I. N. Putri, P. Maharani, Y. E. Kurniawati, and R. A. Putri, "Application of data mining to predict student learning outcomes in Padang Panjang," in *Proc. Int. Conf. Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2024, doi: 10.1109/ICECCME62383.2024.10796870.

[5] Hartatik, K. Kusrini, and A. B. Prasetio, "Prediction of student graduation with Naive Bayes algorithm," in *Proc. 5th Int. Conf. Informatics and Computing (ICIC)*, 2020, doi: 10.1109/ICIC50835.2020.9288625.

[6] A. A. M. E. Hussen and A. Saikhu, "Modeling of student graduation prediction using the Naive Bayes classifier algorithm," in *Proc. 3rd Int. Conf. Creative Communication and Innovative Technology (ICCIT)*, 2024, doi: 10.1109/ICCIT62134.2024.10701117.

[7] P. A. Amahan, "Employing Naive Bayes algorithm in the analysis of students academic performances," in *ACM Int. Conf. Proc. Series*, 2024, pp. 58–61, doi: 10.1145/3647722.3647731.

[8] V. Nakhipova, H. I. Bulbul, Y. Kerimbekov, L. Suleimenova, Z. Umarova, and E. Adylbekova, "Integration of collaborative filtering into Naive Bayes method to enhance student performance prediction," *Int. J. Inf. Commun. Technol. Educ.*, vol. 20, no. 1, 2024, doi: 10.4018/IJICTE.352512.

[9] T. N. Viet, H. L. Minh, L. C. Hieu, and T. H. Anh, "The Naïve Bayes algorithm for learning data analytics," *Indian J. Comput. Sci. Eng.*, vol. 12, no. 4, pp. 1038–1043, 2021, doi: 10.21817/indjcse/2021/v12i4/211204191.

[10] Misinem, T. B. Kurniawan, D. A. Dewi, M. Z. Zakaria, and C. M. A. Nazmi, "Leveraging data analytics for student grade prediction: A comparative study of data features," *J. Appl. Data Sci.*, vol. 5, no. 4, pp. 2025–2038, 2024, doi: 10.47738/jads.v5i4.442.

[11] H. Ma, J. Liu, J. Zhang, and J. Huang, "Estimating the compressive strength of cement-based materials with mining waste using support vector machine, decision tree, and random forest models," *Adv. Civ. Eng.*, vol. 2021, no. 1, Art. no. 6629466, 2021.

[12] H. Syahputra and A. Wibowo, "Comparison of support vector machine (SVM) and random forest algorithm for detection of negative content on websites," *J. Ilm. Tek. Elektro Komput. Inform. (JITEKI)*, vol. 9, no. 1, pp. 165–173, 2023.

[13] A. Damayunita, R. S. Fuadi, and C. Juliane, "Comparative analysis of Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) algorithms for classification of heart disease patients," *J. Online Inform.*, vol. 7, no. 2, pp. 219–225, 2022.

[14] M. J. Nayeem, S. Rana, F. Alam, and M. A. Rahman, "Prediction of hepatitis disease using K-nearest neighbors, Naive Bayes, support vector machine, multi-layer perceptron and random forest," in *Proc. 2021 Int. Conf. Inf. Commun. Technol. Sustainable Develop. (ICICT4SD)*, Feb. 2021, pp. 280–284.