PENERAPAN ALGORITMA HYBRID C4.5 DAN NAÏVE BAYES UNTUK MENENTUKAN JURUSAN KULIAH SISWA DI SMA NEGERI 8 MEDAN BERBASIS WEB

SKRIPSI

DISUSUN OLEH

TIARA LUBIS
NPM. 2109010020

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025

PENERAPAN ALGORITMA HYBRID C4.5 DAN NAÏVE BAYES UNTUK MENENTUKAN JURUSAN KULIAH SISWA DI SMA NEGERI 8 MEDAN BERBASIS WEB

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem Informasi (S.Si) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

`

TIARA LUBIS
NPM. 2109010020

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025

LEMBAR PENGESAHAN

Judul Skripsi	:	PENERAPAN ALGORITMA HYBRID C4.5 DAN NAÏVE BAYES UNTUK MENENTUKAN JURUSAN KULIAH SISWA DI SMA NEGERI 8 MEDAN BERBASIS WEB
Nama Mahasiswa	:	TIARA LUBIS
NPM	:	2109010020
Program Studi	:	SISTEM INFORMASI

Menyetujui
Komisi Pembimbing

(Ferdy Riza ST. M.Kom)
NIDN. 010306891

Ketua Program Studi

(Dr. Firahmi Rizky, S.Kom.,M.Kom)
NIDN. 0129117901

Dekan

(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

ii

PERNYATAAN ORISINALITAS

PENERAPAN ALGORITMA HYBRID C4.5 DAN NAÏVE BAYES UNTUK MENENTUKAN JURUSAN KULIAH SISWA DI SMA NEGERI 8 MEDAN BERBASIS WEB

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, September 2025

Yang membuat pernyataan

Tiara Lubis

NPM. 2109010020

iii

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Tiara Lubis
NPM	: 2109010020
Program Studi	: Sistem Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif *(Non-Exclusive Royalty free Right)* atas penelitian skripsi saya yang berjudul:

PENERAPAN ALGORITMA HYBRID C4.5 DAN NAÏVE BAYES UNTUK MENENTUKAN JURUSAN KULIAH SISWA DI SMA NEGERI 8 MEDAN BERBASIS WEB

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, September 2025

Yang membuat pernyataan

Tiara Lubis

NPM. 2109010020

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Tiara Lubis
Tempat dan Tanggal Lahir	: Medan, 17 Mei 2003
Alamat Rumah	: Jl Rawa Gg Abadi No 16A
Telepon/Faks/HP	: 081378936906
E-mail	: tiarralubiis2@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: SD Negeri 060908	TAMAT: 2015
SMP	: SMP Swasta Nurul Islam Indonesia	TAMAT: 2018
SMA	: SMA Negeri 8 Medan	TAMAT: 2021

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas rahmat, taufik dan karunianya penulis dapat menyelesaikan skripsi yang berjudul "**Penerapan Algoritma Hybrid C4.5 Dan Naive Bayes Untuk Menentukan Jurusan Kuliah Siswa Di SMA Negeri 8 Medan Berbasis**". Shalawat serta salam senantiasa tercurah kepada Nabi Muhammad SAW, yang telah membawa kita dari zaman kegelapan menuju zaman yang terang benderang dengan ilmu pengetahuan.

Skripsi ini ditulis dalam rangka memenuhi syarat untuk mencapai gelar Sarjana Komputer pada Program Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi di Universitas Muhammadiyah Sumatera Utara. Penulis menyadari bahwa penyusunan skripsi ini tidak akan terselesaikan tanpa bantuan dan dukungan dari berbagai pihak. Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., selaku Rektor Universitas Muhammadiyah Sumatera Utara.
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firahmi Rizky S.Kom., M.Kom selaku Ketua Program Studi Sistem Informasi.
4. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom selaku Sekretari Program Studi Sistem Informasi.
5. Bapak Ferdy Riza, ST, M.Kom selaku dosen pembimbing yang telah memberikan arahan dan bimbingan kepada penulis.
6. Kedua orang tua tercinta, yaitu Ummi Ayu Hanna dan Ayahanda

Abdul Rajab, penulis menyampaikan terima kasih yang tak terhingga atas segala kasih sayang yang diberikan dalam membesarkan dan membimbing penulis selama ini. Pengorbanan yang tak kenal lelah, serta memberikan cinta dan doa yang tiada henti dari mereka sangat berarti bagi penulis. Terima kasih juga kepada saudara-saudara penulis yang selalu memberi dukungan serta semangat.

7. Kepada Muhammad Iqbal yang senantiasa mendengarkan keluh kesah saya selama ini, serta memberi dukungan, motivasi, rasa nyaman, kasih sayang dan menemani saya sampai sekarang dan sampai skripsi ini terselesaikan dengan baik.
8. Para rekan seperjuangan Geng Blekping di angkatan 2021 yaitu Maratul Hasanah Via Ningrum, Elca Popi Amada, dan Salsabila Humairoh yang selalu menjadikan setiap perjalanan perkuliahan penuh kenangan. Canda dan kebersamaan kalian bukan sekedar pelengkap hari, tapi pengingat bahwa kuliah tak selalu harus dijalani dengan beban sendiri.
9. Dan seluruh pihak keluarga saya yang selalu memberi doa dan menyayangi penulis dengan tulus. Memberikan semangat, demi kelancaran dan keberhasilan penyusunan skripsi ini.
10. Terakhir tidak lupa, kepada diri saya sendiri. Meskipun memiliki latar belakang tidak sempurna, terimakasih "Araa" sudah memilih untuk bertahan, mau berjuang untuk tetap ada hingga saat ini, serta menjadi perempuan kuat dan ikhlas atas segala perjalanan hidup yang mengecewakan dan menyakitkan itu. Dengan adanya skripsi ini, telah berhasil membuktikan bahwa kamu bisa menyandang gelar S.Kom tepat waktu dan menjadi tekad maupun acuan untuk terus

melakukan hal lebih membanggakan lainnya.

PENERAPAN ALGORITMA HYBRID C4.5 DAN NAÏVE BAYES UNTUK MENENTUKAN JURUSAN KULIAH SISWA DI SMA NEGERI 8 MEDAN BERBASIS WEB

ABSTRAK

Penelitian ini merancang dan membangun sistem rekomendasi jurusan kuliah bagi siswa SMA berbasis web dengan pendekatan hybrid C4.5–Naïve Bayes. Algoritma C4.5 digunakan untuk seleksi fitur dari data nilai rapor dan hasil tes minat, sedangkan Naïve Bayes melakukan klasifikasi jurusan berdasarkan fitur terpilih. Arsitektur sistem memisahkan antarmuka PHP dari layanan analitik Python melalui API, sehingga proses pengolahan data, pelatihan, evaluasi (akurasi, precision, recall, F1), dan penyimpanan model dapat direplikasi di lingkungan lokal. Hasil pengujian skenario end-to-end menunjukkan seluruh alur—mulai dari pengelolaan data, pelatihan dan aktivasi model, pembuatan rekomendasi per siswa/massal, hingga verifikasi oleh wali kelas—berjalan sesuai rancangan, sehingga sistem layak dijadikan alat bantu pengambilan keputusan penjurusan yang objektif dan terukur.

Kata Kunci: C4.5, Naïve Bayes, hybrid, rekomendasi jurusan, sistem pendukung keputusan, SMA, web.

WEB-BASED APPLICATION OF HYBRID C4.5 AND NAÏVE BAYES ALGORITHM TO DETERMINE STUDENTS' COLLEGE MAJORS AT STATE SENIOR HIGH SCHOOL 8 MEDAN

ABSTRACT

This study develops a web-based major recommendation system for high-school students using a hybrid C4.5–Naïve Bayes approach. C4.5 performs feature selection on academic grades and interest-questionnaire data, while Naïve Bayes conducts the classification using the selected features. The system architecture decouples a PHP front end from a Python analytics service via API, enabling local, reproducible workflows for data processing, model training, evaluation (accuracy, precision, recall, F1), and model persistence. End-to-end testing confirms that the complete pipeline—data management, model training and activation, individual/batch recommendations, and counselor verification—operates as designed, indicating the system's suitability as an objective, data-driven decision support tool for student major selection.

Keywords: : C4.5, Naïve Bayes, hybrid, major recommendation, decision support system, high school, web.

DAFTAR ISI

LEMBAR PENGESAHAN ii
PERNYATAAN ORISINALITAS ..... **Error! Bookmark not defined.**
PERNYATAAN PERSETUJUAN PUBLIKASI ..... **Error! Bookmark not defined.**
RIWAYAT HIDUP v
KATA PENGANTAR vi
ABSTRAK viii
ABSTRACT ix
DAFTAR ISI x
DAFTAR GAMBAR xii
DAFTAR TABEL xiv
BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 3
1.5 Manfaat Penelitian 4
BAB II LANDASAN TEORI 5
2.1 Sistem Pendukung Keputusan 5
2.2 Algoritma Naïve Bayes 10
2.3 *Decision Tree* dan Algoritma C4.5 12
2.4 Evaluasi Model Klasifikasi 17
2.5 Teknologi Yang Digunakan *(Platform)* 18
2.6 Flowchart 20
2.7 Penelitian Terdahulu 28
BAB III METOLODOGI PENELITIAN 33
3.1 Jenis Penelitian 33
3.2 Lokasi Penelitian 33
3.3 Tahapan Penelitian 33
3.4 Data dan Sumber Data 36
3.5 Teknik Pengumpulan Data 37
3.6 Teknik Pengolahan dan Analisis Data 37
3.7 Perancangan Sistem 41

3.8 Perancangan Basis Data 65
BAB IV HASIL DAN PEMBAHASAN 77
4.1. Perhitungan Manual 77
4.2. Implementasi Basis Web 91
4.3. Pengujian Program 108
4.4. Kelebihan dan Kelemahan Program 113
BAB V KESIMPULAN DAN SARAN 115
5.1 Kesimpulan 115
5.2 Saran 116
DAFTAR PUSTAKA 116

DAFTAR GAMBAR

Gambar 3. 1 Flowchart Penelitian 33
Gambar 3. 2 Use Case Diagram 43
Gambar 3. 3 Activity Diagram Login 45
Gambar 3. 4 Activity Diagram Dashboard 46
Gambar 3. 5 Activity Diagram Kelola Akun Siswa 47
Gambar 3. 6 Activity Diagram Kelola Mapel 48
Gambar 3. 7 Activity Diagram Kelola Pertanyaan 49
Gambar 3. 8 Activity Diagram Kelola Data Latih 50
Gambar 3. 9 Activity Diagram Analisis dan Latih Model 51
Gambar 3. 10 Activity Diagram Kelola Data Model 52
Gambar 3. 11 Activity Diagram Hasil Rekomendasi 53
Gambar 3. 12 Activity Diagram Verifikasi Hasil 54
Gambar 3. 13 Activity Diagram Logout 55
Gambar 3. 14 Sequence Diagram Login 55
Gambar 3. 15 Sequence Diagram Dashboard 56
Gambar 3. 16 Sequence Diagram Kelola Akun Siswa 57
Gambar 3. 17 Sequence Diagram Kelola Mata Pelajaran 58
Gambar 3. 18 Sequence Diagram Kelola Data Pertanyaan 59
Gambar 3. 19 Sequence Diagram Kelola Data Latih 60
Gambar 3. 20 Sequence Diagram Analisis dan Latih Model 61
Gambar 3. 21 Sequence Diagram Manajemen Data Siswa 62
Gambar 3. 22 Sequence Diagram Hasil Rekomendasi 63
Gambar 3. 23 Sequence Diagram Verifikasi Hasil 64
Gambar 3. 24 Sequence Diagram Logout 64
Gambar 3. 25 Rancangan Class Diagram 65
Gambar 3. 26 Rancangan Tabel Rekomendasi 69
Gambar 3. 27 Interface Menu Login 70
Gambar 3. 28 Interface Dashboard Admin dan Wali Kelas 71
Gambar 3. 29 Interface Kelola Nilai Rapor 71
Gambar 3. 30 Interface Menu Kelola Data Tes Minat 72
Gambar 3. 31 Interface Menu Proses Data dan Latih Model Prediksi 73

Gambar 3. 32 Interface Menu Hasil Rekomendasi.....73
Gambar 3. 33 Interface Menu Verifikasi Hasil 74
Gambar 3. 34 Interface Menu Manajemen Data Siswa.....74
Gambar 3. 35 Interface Lihat Hasil 75
Gambar 4. 1 Tampilan Halaman Login.....88
Gambar 4. 2 Tampilan Halaman Dashboard (Wali Kelas) 89
Gambar 4. 3 Tampilan Halaman Dashboard (Siswa).....89
Gambar 4. 4 Tampilan Halaman Kelola Akun Siswa 90
Gambar 4. 5 Tampilan Halaman Tambah & Edit Siswa.....90
Gambar 4. 6 Tampilan Halaman Data Mata Pelajaran.....91
Gambar 4. 7 Tampilan Tambah & Edit Mata Pelajaran.....92
Gambar 4. 8 Tampilan Halaman Kelola Pertanyaan 92
Gambar 4. 9 Tampilan Halaman Tambah & Edit Pertanyaan.....93
Gambar 4. 10 Tampilan Halaman Kelola Data Latih.....94
Gambar 4. 11 Tampilan Halaman Kelola Data Latih.....95
Gambar 4. 12 Memulai Proses Analisis 96
Gambar 4. 13 Pra Pemrosesan dan Analisis Data Awal.....97
Gambar 4. 14 Seleksi Fitur dengan C4.5.....98
Gambar 4. 15 Pelatihan Naive Bayes - Perhitungan Prior 99
Gambar 4. 16 Pelatihan *Naive Bayes* - Perhitungan *Likelihood*.....99
Gambar 4. 17 Pelatihan Naive Bayes - Perhitungan Posterior.....99
Gambar 4. 18 Evaluasi Performa Model 100
Gambar 4. 19 Tampilan Halaman Kelola Model 101
Gambar 4. 20 Tampilan Hasil Rekomendasi.....102
Gambar 4. 21 Klasifikasi 1 Siswa 103
Gambar 4. 22 Tampilan Klasifikasi Semua Siswa.....104
Gambar 4. 23 Klasifikasi Siswa Baru 105
Gambar 4. 24 Tampilan Hasil Rekomendasi (Siswa) 106
Gambar 4. 25 Tampilan Kelola Verifikasi Hasil.....106
Gambar 4. 26 Tampilan Detail Verifikasi 107

DAFTAR TABEL

Tabel 2. 1 Simbol Flowchart 24
Tabel 2. 2 Use Case Diagram 26
Tabel 2. 3 Simbol Activity Diagram 27
Tabel 2. 4 Simbol Sequence Diagram 28
Tabel 2. 5 Class Diagram 29
Tabel 2. 6 Penelitian Terdahulu 30
Tabel 3. 1 Tahapan Pengumpulan Data 37
Tabel 3. 2 Rancangan Tabel Users 66
Tabel 3. 3 Rancangan Tabel Mata Pelajaran 67
Tabel 3. 4 Rancangan Tabel Nilai Rapor 67
Tabel 3. 5 Rancangan Tabel Pertanyaan Minat 67
Tabel 3. 6 Rancangan Pilihan Jawaban 68
Tabel 3. 7 Rancangan Tabel Jawaban Minat 68
Tabel 3. 8 Rancangan Tabel Model Prediksi 69
Tabel 4. 1 Sampel Data Latih Awal (15 Siswa) 78
Tabel 4. 2 Data Latih Hasil Diskretisasi 78
Tabel 4. 3 3 Atribut dengan Gain Tertinggi 109

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Penentuan jurusan kuliah merupakan salah satu keputusan penting yang dapat memengaruhi masa depan siswa. Namun, proses penjurusan seringkali masih dilakukan secara subjektif, berdasarkan opini atau masukan dari orang tua dan guru, tanpa pemanfaatan data secara objektif (Saputra et al., 2021). Hal ini menyebabkan banyak siswa yang belum memiliki gambaran jelas mengenai jurusan yang sesuai dengan minat, potensi, dan kemampuan akademik mereka, sehingga berisiko mengalami ketidakcocokan jurusan, penurunan motivasi, bahkan keinginan untuk berpindah jurusan (Mustofa & Kurniawan, 2024).

Masalah ini semakin rumit karena konsultasi secara konvensional (manual) yang bersifat subjektif seringkali tidak cukup untuk membantu siswa mengenali kecocokan antara diri mereka dan jurusan yang mereka pilih di perguruan tinggi. Bahkan di tingkat SMA, penjurusan akademik pun masih sering didasarkan pada nilai rapor tanpa analisa lanjutan terhadap kecenderungan minat jurusan yang mau dipilih sisw (Yudhi Putra & Ismiyana Putri, 2023). Tanpa menggunakan metode yang tepat, pola keterkaitan antara atribut nilai dan minat siswa tidak dapat dikenali dengan jelas, sehingga keputusan yang dihasilkan menjadi kurang akurat.

Untuk mengatasi permasalahan tersebut, dibutuhkan sebuah sistem berbasis data yang dapat menentukan jurusan kuliah kepada siswa secara tepat dan sesuai dengan minat mereka secara individual. Sistem ini harus mampu memanfaatkan data yang sudah tersedia, seperti nilai akademik, minat, dan atribut lainnya, untuk memprediksi kecocokan jurusan secara objektif. Nilai akademik mencakup hasil yang diperoleh siswa dalam berbagai mata pelajaran yang relevan, sedangkan minat merujuk pada bidang atau program studi yang diminati siswa, yang dapat dilihat dari preferensi dan kecenderungan mereka terhadap mata pelajaran atau kegiatan tertentu.. Teknologi *Machine Learning*,

khususnya metode klasifikasi telah banyak digunakan untuk membangun sistem prediksi yang lebih akurat dan adaptif (Alharbi & Allohibi, 2024). Dengan adanya sistem ini, siswa dapat memperoleh gambaran mengenai jurusan kuliah yang paling sesuai dengan profil mereka, sehingga proses pengambilan keputusan menjadi lebih terarah dan berdasarkan data.

Dari berbagai metode klasifikasi yang ada, pendekatan *hybrid* antara Naïve Bayes dan C4.5 Decision Tree dapat menjadi pendekatan yang efektif dalam penelitian ini. Naïve Bayes dikenal karena kesederhanaan, efisiensi, serta kemampuannya menangani dataset besar, meskipun memiliki kelemahan pada asumsi independensi antar fitur (Li & Chen, 2024). Sementara itu, C4.5 Decision Tree mampu melakukan seleksi fitur data secara otomatis berdasarkan *information gain*, serta menghasilkan model klasifikasi yang mudah diinterpretasikan (Rim & Liu, 2020). Pendekatan *hybrid* ini mampu meningkatkan kinerja prediksi secara signifikan, memanfaatkan kekuatan seleksi fitur C4.5 dan probabilitas prediksi *Naïve Bayes* (Changpetch et al., 2021).

Dalam penelitian ini, sistem rekomendasi jurusan akan dikembangkan dengan menerapkan pendekatan *hybrid Naïve Bayes* dan C4.5. Data yang digunakan meliputi nilai akademik siswa, serta data minat studi yang diperoleh dari pihak sekolah. Proses seleksi fitur dan pengelompokkan awal dilakukan menggunakan C4.5, kemudian hasilnya digunakan sebagai input bagi model *Naïve Bayes* untuk melakukan klasifikasi jurusan yang penting bagi setiap siswa.

Sistem yang dikembangkan berbasis web menggunakan Javascript dan perangkat tambahan untuk analisis data lainnya, dengan tujuan untuk mempermudah proses visualisasi serta evaluasi akurasi dari hasil rekomendasi (Changpetch et al., 2021). Dengan rekomendasi yang menggabungkan pendekatan hibrida Naïve Bayes dan C4.5 ini, diharapkan siswa memperoleh saran jurusan kuliah yang lebih tepat, sesuai dengan minat, kemampuan, dan potensi yang dimiliki. Sistem ini juga berfungsi untuk membantu siswa dalam membuat keputusan yang lebih percaya diri saat memilih jurusan, sehingga keputusan yang diambil menjadi lebih matang dan selaras dengan karakteristik individu. Oleh karena itu, skripsi ini berjudul "Penerapan Algoritma C4.5 dan

Naïve Bayes untuk Menentukan Jurusan Kuliah pada Siswa SMA Negeri 8 Medan berbasis Web."

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana merancang dan mengembangkan sistem berbasis web untuk penentuan jurusan kuliah dengan metode *hybrid* C4.5 dan *Naïve Bayes*?
2. Seberapa efektif kinerja model hybrid Naïve Bayes dan C4.5 dalam memberikan jurusan yang tepat berdasarkan metrik akurasi, *precision, recall,* dan *F1-Score*?

1.3 Batasan Masalah

Untuk menjaga ruang lingkup penelitian agar tetap fokus dan terarah, maka Batasan masalah dalam penelitian ini ditetapkan sebagai berikut :

1. Penelitian ini hanya membahas metode *hybrid* C4.5 dan *Naïve Bayes*, tanpa membandingkannya dengan metode lain.
2. Sistem yang dibuat adalah *prototype,* dan belum jadi yang siap dipakai di sekolah. Untuk tampilan *front end* menggunakan PHP, untuk analisis data dan penggabungan kedua metode menggunakan *Javascript*
3. Data yang diolah hanya nilai rapor dan data minat siswa. Faktor lain seperti prestasi non-akademik disertakan jika disediakan oleh sekolah.
4. Pengujian hanya berfokus pada fungsi dasar dan efektifitas dari sistem yang dibuat.
5. Sistem ini hanya memberikan saran jurusan sebagai bahan pertimbangan, bukan alat untuk pendaftaran kuliah.

1.4 Tujuan Penelitian

Adapun yang menjadi tujuan dalam penelitian ini adalah sebagai berikut :

1. Merancang dan mengembangkan sistem berbasis web untuk penentuan

jurusan kuliah yang mengimplementasikan metode hybrid C4.5 dan Naïve Bayes.

2. Menerapkan metode hybrid C4.5 dan Naïve Bayes untuk menghasilkan rekomendasi jurusan kuliah yang akurat bagi siswa, dengan mengintegrasikan algoritma C4.5 untuk seleksi fitur dan algoritma Naïve Bayes untuk klasifikasi.

1.5 Manfaat Penelitian

Adapun yang menjadi manfaat dalam penelitian ini adalah sebagai berikut

1. Menyediakan sistem rekomendasi jurusan kuliah berbasis data yang dapat membantu siswa dalam memilih jurusan secara lebih tepat dan objektif, dengan memperhitungkan nilai rapor dan minat siswa.
2. Memberikan kontribusi dalam pengembangan aplikasi berbasis metode hybrid C4.5 dan Naïve Bayes dalam bidang pendidikan, khususnya untuk sistem rekomendasi yang dapat digunakan oleh sekolah dalam menentukan jurusan siswa.
3. Membantu sekolah, guru, dan siswa dalam pengambilan keputusan penjurusan yang lebih efisien, terarah, dan berbasis data, serta membuka peluang untuk pengembangan sistem pendukung keputusan lebih lanjut.

BAB II
LANDASAN TEORI

2.1 Sistem Pendukung Keputusan

Sistem Pendukung Keputusan (SPK) atau *Decision Support System* (DSS) adalah sistem informasi berbasis komputer yang dirancang untuk membantu pengambil keputusan dalam menghadapi masalah yang bersifat semi-terstruktur atau tidak terstruktur. SPK mengintegrasikan data, model analitis, dan antarmuka pengguna untuk mendukung proses pengambilan keputusan yang lebih efektif dan efisien. Menurut penelitian oleh Putra dan Nasir (2021), SPK dirancang untuk menunjang seluruh tahapan pembuatan keputusan, mulai dari identifikasi masalah hingga evaluasi alternatif solusi.

Sebuah SPK umumnya terdiri dari tiga komponen utama:

1. **Subsistem Manajemen Data**: Berfungsi sebagai penyimpan, pengelola, dan penyedia data yang relevan dalam proses pengambilan keputusan. Data dapat berasal dari berbagai sumber, baik internal maupun eksternal organisasi .
2. **Subsistem Manajemen Model**: Menyediakan berbagai model analitis atau algoritma yang digunakan untuk mengolah data dan menghasilkan informasi yang mendukung keputusan. Model ini bisa berupa model matematis, statistik, atau kecerdasan buatan.
3. **Subsistem Antarmuka Pengguna** : Menyediakan sarana interaksi antara pengguna dan sistem, memungkinkan pengguna untuk memasukkan data, menjalankan model, dan melihat hasil analisis dengan cara yang mudah dipahami (Unesa, 2025).

Dalam konteks bisnis, Sistem pendukung Keputusan dapat memberikan manfaat sebagai berikut :

1. **Meningkatkan Efektivitas Keputusan**: Dengan menyediakan data dan analisis yang akurat, SPK membantu manajer membuat keputusan yang lebih tepat dan berbasis informasi (Risa Satifa, 2023).
2. **Mengidentifikasi Masalah dan Peluang**: SPK memungkinkan

perusahaan untuk mengenali masalah potensial dan peluang bisnis melalui analisis data yang mendalam.

3. **Mengoptimalkan Proses Bisnis**: Dengan analisis yang tepat, SPK dapat membantu dalam perencanaan sumber daya, manajemen rantai pasok, dan strategi pemasaran yang lebih efektif (Unesa, 2025).

Proses pengambilan keputusan yang didukung oleh SPK umumnya mengikuti serangkaian tahapan yang sistematis. Simon (1960) mempopulerkan model empat fase yang menjadi kerangka kerja dasar bagi banyak SPK.

1. Fase Intelijen *(Intelligence Phase)*

   Ini adalah tahap awal di mana masalah diidentifikasi dan didefinisikan. Aktivitas utamanya adalah penelusuran lingkungan untuk mengenali situasi atau problematika yang memerlukan keputusan. Pada fase ini, data mentah dari berbagai sumber (internal dan eksternal) dikumpulkan, diproses, dan dianalisis untuk mengidentifikasi akar masalah, kendala, dan peluang yang ada. SPK membantu dengan menyediakan laporan, kueri, dan visualisasi data yang relevan.

2. Fase Desain *(*Design *Phase)*

   Setelah masalah dipahami, tahap selanjutnya adalah merancang dan mengembangkan berbagai alternatif solusi yang mungkin. Pada fase ini, SPK membantu manajer dalam membangun model analisis, seperti model statistik atau kuantitatif, untuk mengeksplorasi berbagai skenario ("what-if analysis"). Proses ini melibatkan verifikasi dan validasi untuk memastikan bahwa model yang digunakan akurat dan relevan dengan masalah yang dihadapi. Tujuannya adalah menghasilkan serangkaian solusi yang layak untuk dievaluasi lebih lanjut.

3. Fase Pilihan *(*Choice *Phase)*

   Pada tahap ini, dilakukan evaluasi terhadap setiap alternatif solusi yang telah dikembangkan pada fase desain. Tujuannya adalah memilih satu alternatif terbaik berdasarkan kriteria dan tujuan yang telah ditetapkan. SPK sangat berperan dalam fase ini dengan menyediakan alat untuk membandingkan alternatif secara objektif, menganalisis sensitivitas, dan

memberikan peringkat. Hasil analisis ini menjadi dasar bagi manajer untuk membuat keputusan akhir.

4. Fase *Implementasi (Implementation Phase)*

Tahap terakhir adalah menerapkan solusi yang telah dipilih dan memantau hasilnya. SPK dapat membantu dalam fase ini dengan menyediakan laporan kemajuan dan analisis kinerja untuk memastikan bahwa solusi yang diimplementasikan berjalan sesuai rencana. Jika ditemukan penyimpangan, informasi ini dapat menjadi masukan untuk kembali ke fase intelijen dan memulai siklus pengambilan keputusan yang baru.

Sistem Pendukung Keputusan (SPK) adalah sistem yang komprehensif, terdiri dari tiga komponen utama yang saling berinteraksi untuk menyediakan fungsionalitas dalam mendukung pengambilan keputusan. Ketiga komponen ini bekerja secara sinergis untuk mengelola data, model, dan interaksi dengan pengguna.

1. Subsistem Manajemen Data (*Data Management Subsystem*)

Subsistem ini berfungsi sebagai fondasi dari SPK karena berisi semua data yang relevan untuk proses pengambilan keputusan. Komponen-komponen dalam subsistem ini meliputi:

a. Basis Data (Database): Kumpulan data yang berasal dari sumber internal (misalnya, data transaksi perusahaan) dan eksternal (misalnya, data pasar).
b. DBMS (*Database Management System*): Perangkat lunak yang mengelola dan mengatur data dalam basis data, memungkinkan ekstraksi, penambahan, dan manipulasi data.
c. Koneksi ke *Data Warehouse* : Terhubung dengan data warehouse, yaitu repositori data perusahaan yang terintegrasi dan relevan untuk analisis keputusan.
d. Kemampuan utama dari subsistem ini adalah mengelola berbagai jenis data dan menggabungkannya untuk analisis yang lebih lanjut.

2. Subsistem Manajemen Model (*Model Management Subsystem*)

Komponen ini merupakan "otak" dari SPK, berisi berbagai model kuantitatif yang digunakan untuk menganalisis data dan menghasilkan alternatif keputusan. Beberapa komponen yang ada dalam subsistem ini adalah:

a. Model Finansial, Statistik, dan Kuantitatif: Termasuk model optimisasi, simulasi, dan peramalan.
b. MBMS (*Model Base Management System*): Perangkat lunak yang digunakan untuk membuat, menyimpan, mengakses, dan mengintegrasikan model-model yang ada.
c. Keunggulan SPK terletak pada kemampuannya untuk mengintegrasikan akses data dengan model-model keputusan, memungkinkan analisis yang lebih kompleks dan mendalam.

3. Subsistem Dialog (*User Interface Subsystem*)

a. Komponen ini berfungsi sebagai jembatan antara pengguna (manajer) dan sistem. Antarmuka pengguna yang baik sangat penting untuk memastikan SPK mudah digunakan dan interaktif. Subsistem ini meliputi:
b. Bahasa Aksi: Cara pengguna berkomunikasi dengan sistem (misalnya, melalui menu, perintah, atau formulir).
c. Bahasa Tampilan: Cara sistem menyajikan hasil kepada pengguna (misalnya, dalam bentuk grafik, laporan, atau tabel).
d. Basis Pengetahuan: Informasi yang dibutuhkan pengguna untuk memahami dan menggunakan sistem secara efektif.
e. Fleksibilitas antarmuka ini memungkinkan pengguna dari berbagai tingkatan untuk berinteraksi dengan sistem sesuai kebutuhan mereka.

Sistem Pendukung Keputusan (SPK) memiliki serangkaian karakteristik yang membedakannya dari sistem informasi lainnya, di antaranya:

1. Dukungan untuk Masalah Semi-Terstruktur & Tak Terstruktur

   Fokus utama SPK adalah membantu mengatasi masalah yang tidak memiliki solusi pasti, yaitu masalah yang kompleks dan tidak terstruktur.

2. Dukungan untuk Semua Level Manajerial

   SPK dapat digunakan oleh manajer di semua tingkatan, mulai dari manajer operasional hingga manajer strategis.

3. Dukungan untuk Individu dan Kelompok

   SPK mendukung pengambilan keputusan baik secara individu maupun kolaboratif dalam sebuah tim.

4. Adaptif dan Fleksibel

   SPK dapat beradaptasi dengan perubahan kondisi dan kebutuhan pengguna seiring waktu.

5. Interaktif dan Mudah Digunakan

   SPK dirancang agar pengguna merasa nyaman saat berinteraksi dengan sistem, yang sering kali dilakukan melalui antarmuka grafis yang intuitif.

6. Fokus pada Efektivitas

   Tujuan utama SPK adalah meningkatkan kualitas keputusan yang diambil, bukan hanya sekadar efisiensi dalam proses.

7. Manusia Mengontrol Mesin

   Pengambil keputusan tetap memiliki kontrol penuh atas proses pengambilan keputusan. SPK hanya berfungsi sebagai alat pendukung, bukan pengganti manusia.

8. Berbasis Model dan Data

   SPK menggunakan model untuk menganalisis data yang berasal dari berbagai sumber untuk menghasilkan keputusan yang optimal

2.2 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan salah satu algoritma klasifikasi yang populer. Algoritma ini didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur (features) yang disebut sebagai "naïve" (naif). Meskipun membuat asumsi yang kuat mengenai independensi, algoritma ini dapat bekerja dengan sangat baik dalam berbagai kasus sistem seperti klasifikasi teks, analisis sentimen, dan prediksi diagnosis penyakit, bahkan dengan data yang terbatas. Keunggulan utama dari algoritma Naïve Bayes terletak pada kesederhanaan, efisiensi komputasi, serta kemampuannya untuk menghasilkan model yang baik meskipun dengan sedikit data pelatihan (Chen et al., 2021).

2.2.1. Prinsip Kerja Algoritma Naïve Bayes

Prinsip dasar algoritma Naïve Bayes adalah menghitung probabilitas posterior dari suatu kelas (misalnya, kelas prediksi untuk jurusan kuliah) diberikan data fitur X. Teorema Bayes menyatakan bahwa probabilitas posterior P(H|X) dapat dihitung dengan rumus berikut:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad \text{.....(2.1)}$$

Dimana,

1. $P(H|X)$ adalah probabilitas posterior, yaitu probabilitas bahwa hipotesis H benar setelah kita mengamati data X (misalnya, prediksi jurusan mahasiswa berdasarkan profilnya).

2. $P(X|H)$ adalah *likelihood*, yaitu probabilitas bahwa data X akan diamati jika hipotesis H benar.

3. $P(H)$ adalah probabilitas prior, yaitu probabilitas awal dari kelas H sebelum kita mengamati data X.

4. $P(X)$ adalah normalisasi, yang sering dianggap konstan karena kita hanya tertarik pada kelas dengan probabilitas posterior tertinggi, sehingga dapat diabaikan dalam perhitungan.

2.2.2. Asumsi Independensi Antar Fitur

Pada algoritma Naïve Bayes, terdapat asumsi bahwa fitur-fitur dalam data saling independen satu sama lain. Ini berarti bahwa, meskipun fitur-fitur tersebut bisa saling berhubungan dalam kenyataannya, algoritma ini menganggap bahwa semua fitur tidak saling mempengaruhi satu sama lain.

Jika X = (x1, x2, ….., xn) adalah sekumpulan fitur dalam data, maka *likelihood* P(X|H) dapat dihitung sebagai produk dari probabilitas masing-masing fitur :

$$P(X|H)=P(x1|H)\cdot P(x2|H)\cdot ...\cdot P(xn|H) \quad (2.2)$$

2.2.3. Perhitungan Probabilitas Posterior untuk Klasifikasi

Langkah berikutnya adalah menghitung probabilitas posterior untuk setiap kelas H1, H2, ……, Hk Setelah menghitung likelihood untuk setiap fitur dan probabilitas prior untuk setiap kelas, kita dapat menghitung probabilitas posterior untuk masing-masing kelas dengan menggunakan rumus Teorema Bayes:

$$P(H|X)=\frac{P(X|Hi)P(Hi)}{P(x)} \quad (2.3)$$

i

2.2.4. Perhitungan Probabilitas *Likelihood*

Untuk setiap kelas H_i , dihitung probabilitas *likelihood* P(X | Hi). Menggunakan asumsi independensi antar fitur, kita dapat menghitungnya sebagai produk dari probabilitas untuk setiap fitur seperti pada rumus 2.2.2

2.2.5. Perhitungan Probabilitas Prior

Probabilitas prior P(Hi) adalah probabilitas awal bahwa data termasuk dalam kelas Hi sebelum mengamati data X. Probabilitas prior ini dihitung dengan cara membagi jumlah data dalam kelas Hi dengan total data:

$$P(H_i) = \frac{N(H_i)}{N(D)} \quad (2.4)$$

Dimana :

$N(H_i)$ adalah jumlah data dalam kelas

Hi N(D) adalah jumlah data dalam

dataset

2.3 *Decision Tree* dan Algoritma C4.5

Pohon Keputusan *(Decision Tree)* merepresentasikan prosedur pengambilan keputusan melalui serangkaian pengujian fitur hingga mencapai daun (kelas / Keputusan). C4.5 adalah pengembangan dari ID3 yang memperkenalkan *gain ratio*, yang mampu menangani atribut numerik serta nilai hilang, dan tetap relevan di lanskap *modern* karena kecepatan, efisiensi komputasi, serta interpretabilitasnya (Blockeel et al., 2023).

Gelombang riset *decision tree* menguat sejak 1980-an (ID3/CART), lalu disempurnakan menjadi C4.5 dan berkembang ke *forests/ensembles*. Perkembangan mutakhir menempatkan pohon/*forest* pada konteks *Responsible AI* (*fairness, robustness, explainability*) seraya mempertahankan peran pentingnya pada data tabular.

Induksi pohon dilakukan dengan membagi data secara rekursif: memilih uji/atribut paling informatif di setiap simpul, mempartisi data menurut hasil uji, dan mengulang proses pada *subset* sampai kriteria berhenti terpenuhi, kemudian *pruning* untuk mencegah *overfitting*. Strategi pemilihan uji dan pruning beragam, namun tujuannya sama: memperoleh struktur yang ringkas dan

generalisasi yang baik. Adapun rumus-rumus yang umumnya digunakan

1. *Entropy*

Entropy mengukur seberapa acak/tidak murni distribusi kelas/label pada sebuah himpunan data S. Semakin besar entropi, semakin campuir aduk kelasnya, Entropi dipakai sebagai dasar untuk menghitung *Information Gain*.

Adapun rumus *Entropy* adalah sebagai berikut :

$$Entropy\,(S) = -\sum_{c} p(c)\,log_2 p(c)$$

Dimana:

S = Kumpulan data pada *node* atau cabang pohon pada saat ini

c = setiap kelas (misalkan : Rumpun RTI, Soshum, dsb)

p (c) = proporsi data di S yang berlabel kelas c

log2 = Logaritma basis 2 (mengukur "bit" informasi)

Konteks pemakaian rumus *entropy* ini bermanfaat dalam proses pemilihan atribut/akar

2. *Information Gain*

Information Gain mengukur berapa banyak *entropy (S)* berkurang setelah data S dipecah menurut atribut A. Atribut dengan *information gain* terbesar dianggap paling informatif pada *node* tersebut dan dipilih sebagai pemisah. Ini adalah inti dari *recursive partitioning* (pembentukan pohon dari atas kebawah) (Rasekh & Bavil, n.d.). Rumus yang dipakai dalam *Information Gain* adalah sebagai berikut :

$$Gain\,(S, A) = Entropy\,(S) - \sum_{v \in a} \frac{|S_v|}{|S|} Entropy\,(S_v)$$

Dimana :

A = Atribut yang diuji (misalkan Nilai Bahasa Indonesia

$v \in a$ = Setiap nilai/cabang atribut A (misalkan Tinggi, Sedang, Rendah)

S_v = Subset data di S yang memiliki nilai v

$|S_v|$ / |S| = Bobot (proporsi) masing-masing cabang Adapun prosedur perhitungan :

1. Hitung *Entropy (S)* pada simpul saat ini
2. Untuk setiap atribut A, pecah data jadi cabang-cabang berdasarkan nilai A (atau dua sisi ambang bagi atribut numerik), hitung *Entropy* (S_v) dan hitung IG (S,A).
3. Pilih atribut dengan *Information Gain* tertinggi sebagai pemisah simpul, ulangi ke anak-simpul sampai syarat berhenti terpenuhi (semua contoh satu kelas; atau tak ada atribut tersisa; atau cabang kosong).

Pada fitur berbentuk angka (misalnya nilai rapor). C4.5 akan mencari batas/ambang terbaik, dilambangkan t, untuk membagi data jadi dua :

Kelompok 1 : Nilai <= t Kelompok 2 :Nilai > t

Setiap dibagi, kedua kelompok itu punya komposisi kelas yang lebih rapi/bersih (tidak terlalu campur), sehingga entropi turun. Besarnya penurunan entropi itulah *Information Gain* (IG). Ambang terbaik adalah ambang yang menghasilkan IG paling besar.

Langkah sederhana di setiap *node* :

a) Ambil beberapa kandidat ambang (t) dari data (biasanya di titik Tengah antara dua nilai yang berurutan)
b) Untuk setiap t :
   1) Bagi data jadi dua: ≤ t dan > t.
   2) Hitung entropi di masing-masing sisi.
   3) Hitung IG = entropi sebelum dibagi − (rata-rata entropi dua sisi)
c) Pilih t* dengan IG tertinggi untuk fitur tersebut.
d) Ulangi langkah yang sama untuk fitur numerik lain, lalu pilih fitur (beserta t*- nya) dengan IG paling besar sebagai pemisah node.

Adapun ukuran kinerja Algoritma C4.5 dapat menggunakan beberapa

metrik dibawah ini :

a) Akurasi sebagai ukuran utama dan cara pengujiannya

Dalam evaluasi model pohon keputusan, akurasi secara luas dipandang sebagai ukuran kinerja yang paling sesuai/utama. Praktik penilaian biasanya menghitung akurasi pada data uji yang terpisah dari data latih, atau melalui validasi silang (cross-validation), misalnya *k-fold*; contoh yang eksplisit adalah 2-fold cross- validation: data dibagi menjadi dua bagian, dilatih pada satu bagian dan diuji pada bagian lainnya, lalu ditukar dan dirata-ratakan hasilnya. Pendekatan ini memastikan penilaian yang reliabel atas kemampuan generalisasi model (Rasekh & Bavil, n.d.).

2.4 Kombinasi Algoritma *Naïve Bayes* dan C4.5

Penggabungan *(hybrid)* antara algoritma C4.5 dan Naïve Bayes bertujuan untuk memanfaatkan keunggulan masing-masing metode guna menghasilkan model klasifikasi yang lebih akurat. Dalam pendekatan ini, C4.5 tidak digunakan sebagai klasifikasi akhir, melainkan sebagai mekanisme seleksi fitur (*feature selection*). Alur kerja dari metode *hybrid* dapat diuraikan dalam langkah-langkah berikut :

1. Tahap 1 : Pembangunan Pohon Keputusan C4.5

   Langkah pertama adalah membangun model pohon keputusan menggunakan algoritma C4.5 dari keseluruhan data latih. Proses ini melibatkan perhitungan *Information Gain* untuk setiap atribut guna menentukan node akar dan cabang terbaik, sesuai dengan rumus berikut :

$$Entropy(S)_{i\equiv 1} \sum^{n} \quad -pi * log_2 pi \tag{2.7}$$

Dimana :

S : Kumpulan Kasus

N : Jumlah partisi S

Pi : proporsi dari Si terhadap S

2. Tahap 2 : Ekstraksi dan Seleksi Fitur

   Setelah model pohon keputusan C4.5 terbentuk, tahap selanjutnya

adalah proses identifikasi dan ekstraksi fitur yang paling relevan. Fitur-fitur relevan ini merupakan atribut yang oleh algoritma C4.5 ditempatkan sebagai node pemisah (*decision node*) dalam struktur pohon, yang menandakan kontribusi paling signifikan dalam proses klasifikasi.

Atribut-atribut dengan nilai *Information Gain* tertinggi akan dipilih untuk digunakan pada tahap berikutnya, sementara fitur yang dinilai kurang relevan akan diabaikan. Proses ini bertujuan untuk mengurangi *noise* (gangguan) dari data yang tidak berpengaruh signifikan terhadap hasil akhir, sehingga analisis menjadi lebih fokus dan efisien.

3. Tahap 3 : Pelatihan Model *Naïve Bayes*

Selanjutnya, algoritma klasifikasi Naïve Bayes dilatih. Proses pelatihan ini tidak menggunakan keseluruhan himpunan data awal, melainkan hanya memanfaatkan subset fitur yang telah terseleksi melalui algoritma C4.5 pada tahap sebelumnya. Dengan data yang lebih ringkas dan relevan ini, model Naïve Bayes dibangun untuk mempelajari pola probabilitas antara profil siswa dengan kelas jurusan yang telah ditentukan. Tujuan dari tahap ini adalah untuk menghitung probabilitas posterior, yaitu peluang suatu data (X′) masuk ke dalam kelas hipotesis
(H) tertentu. Perhitungan ini didasarkan pada Teorema Bayes sebagai berikut :

$$P(H|X^i) = \frac{P\ (X^i|H)\ .\ P(H)}{P(X^i)} \quad \text{.....(2.8)}$$

Dimana :

H adalah Hipotesis kelas (contoh saintek atau soshum).

X^i adalah data siswa yang hanya berisi fitur-fitur terpilih.

$P(H|X^i)$ adalah adalah probabilitas posterior, yaitu probabilitas bahwa hipotesis H benar setelah data X′ diobservasi.

4. Tahap 4 : Klasifikasi dengan *Naïve Bayes*

Tahap akhir adalah penerapan model Naïve Bayes yang telah dilatih untuk melakukan prediksi jurusan pada data baru. Model akan menghitung

nilai probabilitas untuk setiap kemungkinan kelas jurusan berdasarkan nilai dari fitur- fitur terpilih pada data siswa tersebut. Perhitungan *likelihood* atau P(X′|H) didasarkan pada produk probabilitas dari setiap fitur secara independen , sesuai dengan rumus berikut:

$$P(X^i \mid H) = \prod_{i=1}^{m} P(xi' \mid H)$$

Kelas jurusan dengan nilai probabilitas posterior tertinggi akan ditetapkan sebagai hasil akhir rekomendasi dari sistem. Pendekatan *hybrid* ini secara efektif memanfaatkan kekuatan C4.5 dalam seleksi fitur dan kemampuan prediksi probabilistik dari Naïve Bayes untuk menghasilkan klasifikasi yang akurat.

2.5 Evaluasi Model Klasifikasi

Confussion Matrix adalah sebuah table yang menunjukkan hasil klasifikasi, baik yang benar maupun yang salah (Normawati & Prayogi, 2021). Tabel ini memiliki empat istilah utama

1. *True Negative* (TN) yang berarti model klasifikasi memprediksi data ada di kelas *negative* namun yang sebenarnya data memang ada di kelas *negative*.
2. *True Positive* (TP) yang berarti model klasifikasi memprediksi data ada dikelas positif namun sebenarnya memang data berada di kelas positif.
3. *False Negative* (FN) yang berarti model klasifikasi memprediksi data ada di kelas *negative* namun yang sebenarnya data ada di kelas positif.
4. *False Positive* (FP) yang berarti model klasifikasi memprediksi data ada di kelas positif namun yang sebenarnya data ada di kelas *negative*.

Dari defenisi *confusion matrix*, beberapa poin poin dalam *confusion matrix* digunakan untuk menghitung *precision, recall,* dan *f1 score*. *Precision* adalah perbandingan antara *True Positive* (TP) dengan banyaknya data yang diprediksi positif, secara matematis dapat dilihat

dibawah ini

$$Precision = \frac{TP}{TP+FP} \quad \text{..... (2.7)}$$

Untuk *recall* sendiri adalah perbandingan antara *true positive* (TP) dengan banyaknya data yang sebenarnya positif. Dapat dinyatakan secara matematis seperti
dibawah ini

$$Recall = \frac{\text{TP}}{TP + FN}$$

Sedangkan *F1 Score* adalah nilai tengah dari *precision* dan *recall*. Nilai terbaik dari *F1 Score* adalah 1 dan nilai terburuknya adalah 0, secara matematis dapat dituliskan seperti dibawah ini

$$F1 = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision+Recall}} \quad \text{..... (2.9)}$$

Nilai *F1 Score* yang baik menandakan bahwa model klasifikasi kita punya *precision* dan *recall* yang baik (Setiawan, 2020).

2.6 Teknologi Yang Digunakan *(Platform)*

2.6.1 HTML

HTML, atau *HyperText Markup Language*, adalah bahasa markup standar yang digunakan untuk membuat struktur halaman web. HTML memberikan struktur dasar pada setiap halaman web dengan menggunakan berbagai elemen seperti teks, gambar, tautan, dan form. HTML mendefinisikan elemen-elemen tersebut dalam bentuk tag, yang biasanya dibuka dan ditutup dengan tanda kurung sudut. Setiap tag memiliki fungsi khusus, seperti <h1> untuk heading, <p> untuk paragraf, dan <a> untuk hyperlink. Versi terbaru dari HTML, yaitu HTML5, menambahkan banyak elemen semantik yang mendukung pengelolaan konten multimedia, formulir interaktif, serta pengoptimalan untuk perangkat mobile.

HTML memiliki peran yang sangat penting dalam pengembangan web modern, karena ia memberikan fondasi untuk penataan konten dan navigasi situs web. Keberhasilan penggunaan HTML yang baik dapat meningkatkan pengalaman pengguna dan aksesibilitas web. Sebagai contoh, HTML5 mengintroduksi elemen- elemen baru seperti <article>, <section>, dan <footer> yang membuat struktur halaman web lebih semantik dan mudah diinterpretasikan oleh mesin pencari serta alat bantu seperti pembaca layar. Ini berkontribusi pada peningkatan SEO (Search Engine Optimization) dan memudahkan pengelolaan konten dinamis (Yusuf, 2022).

2.6.2 CSS

CSS adalah bahasa yang digunakan untuk mendesain dan mengatur tampilan visual halaman web, termasuk tata letak, warna, font, dan elemen desain lainnya. Dengan CSS, pengembang dapat menciptakan antarmuka pengguna yang menarik dan responsif.

2.6.3 MySQL

MySQL adalah RDBMS yang cepat dan mudah digunakan, serta sudah banyak dipakai untuk berbagai kebutuhan. MySQL dikembangkan oleh MySQL AB Swedia. Hampir sebagian besar sistem berbasis website yang ada di internet dikembangkan menggunakan MySQL dan bahasa pemrograman lainnya, seperti PHP. Sistem database MySQL mendukung beberapa fitur seperti *multithreaded*, *multiuser* dan *SQL Database Management System* (DBMS) (Naldo et al., 2021).

Menurut Hidayat dkk., (2019) MySQL adalah sebuah implementasi dari sistem manajemen basis data relasional (RDBMS) yang didistribusikan secara gratis dibawah lisensi GPL (*General Public License*). Setiap pengguna dapat secara bebas menggunakan MySQL, namun dengan batasan perangkat lunak tersebut tidak boleh dijadikan produk turunan yang bersifat komersial.

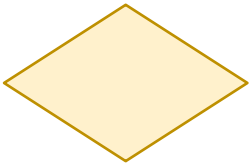
Berikut ini hal-hal yang menyebabkan MySQL menjadi begitu populer (Ari Nurjani et al., 2021):

1. Berlisensi open-source, sehingga kita dapat menggunakannya secara gratis.
2. Merupakan program yang powerful dan menyediakan fitur yang lengkap.
3. Menggunakan bentuk standar bahasa data SQL.
4. Dapat bekerja dengan banyak sistem operasi dan dengan bahasa-bahasa pemrograman, seperti PHP, PERL, C, C++, JAVA, dan lain- lain.
5. Bekerja dengan cepat dan baik, bahkan dengan data set yang banyak.
6. Sangat mudah digunakan dengan PHP untuk pengembangan sistem berbasis web.
7. Mendukung banyak database, sampai 50 juta baris atau lebih dalam suatu tabel.
8. Dapat dikustomisasi sesuai dengan keinginan.

2.7 Flowchart

Flowchart adalah bagan-bagan yang mempunyai arus yang menggambarkan langkahlangkah penyelesaian suatu masalah. penggambaran secara grafik dari langkah-langkah dan uruturutan prosedur dari suatu program. *Flowchart* menolong analis dan programer untuk memecahkan masalah kedalam segmen-segmen yang lebih kecil dan menolong dalam menganalisis alternatif-alternatif lain dalam pengoperasian (khoerul ummah, 2022). *Flowchart* terdiri dari simbol-simbol grafis, *flowchart* mengilustrasikan tugas, keputusan, serta aliran data dalam alur kerja yang bertujuan untuk mempermudah pemahaman, merancang, dan menganalisis agar dapat dipahami dengan mudah. Setiap *step* atau tugas dalam *flowchart* digambarkan dengan menggunakan simbol khusus dan dihubungkan dengan garis atau panah. Berikut merupakan simbol flowchart, nama, serta arti dari tiap simbol.

Tabel 2. 1 Simbol *Flowchart*

No.	Simbol *Flowchart*	Nama	Arti Simbol *Flowchart*
1		*Terminator*	Awalan atau akhir konsep (prosedur)
2		*Process*	Penggambaran bentuk proses operasional
3		*Document*	Dokumen/laporan berbentuk *print out*
4		*Decision*	Keputusan atau *sub-pont.*
5		*Data*	*Input* atau *output*
6		*On-Page Reference/Connector*	Penghubung alur dalam *page* yang sama

7		*Off-Page Reference /Off-Page Connector*	Penghubung alur dalam *different page*
8		*Flow*	*Arah alur dalam sebuah prosedur*

(Sumber : Rumbaugh et al., 2021**).**

Terdapat 8 simbol *flowchart* yang masing-masing memiliki nama dan arti yang berbeda. *Flowchart* memiliki manfaat yang besar dalam membantu mengkomunikasikan ide dan rencana secara visual dan mudah untuk dipahami oleh orang yang melihatnya.

2.7.1 *Unified Modelling Language* (UML)

Unified Modelling Language (UML) adalah bahasa pemodelan visual yang digunakan untuk menentukan, memvisualisasikan, membangun berbagai aspek dari sebuah sistem perangkat lunak. UML sendiri memiliki fungsi sebagai alat untuk menangkap pemahaman mengenai sistem yang perlu dibangun.

Bayangkan sistem sebagai kelompok objek yang berbeda yang bekerja sama untuk melakukan pekerjaan yang berguna untuk pengguna. Bagian struktur statis mendefiniskan jenis objek yang penting untuk sistem dan bagaimana mereka terkait satu sama lain. Bagian perilaku dinamis menggambarkan bagaimana objek berubah seiring waktu dan berkomunikasi satu sama lain untuk mencapai tujuan tertentu. Dengan

memodelkan sistem dari berbagai sudut pandang yang terkait, kita dapat memahami sistem tersebut untuk berbagai keperluan (Rumbaugh et al.., 2021).

2.7.2 Use Case Diagram

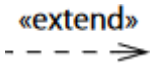
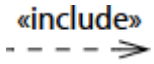
Use Case Diagram menggambarkan bagaimana sistem berperilaku ketika digunakan oleh pengguna eksternal. Diagram ini memecah fungsionalitas sistem menjadi tindakan-tindakan yang bermakna bagi pengguna (Aktor) yang berinteraksi dengan sistem. Aktor bisa mencakup manusia, serta sistem dan proses komputer lainnya.

Aktor merupakan representasi ideal dari orang, proses, atau objek eksternal yang berinteraksi dengan suatu sistem, subsistem, atau kelas. Untuk mengidentifikasi Aktor, perlu ditentukan tugas-tugas yang terkait dengan peran mereka dalam konteks sistem. Seseorang atau sistem bisa memainkan beberapa peran, dan penting untuk mengetahui apakah Aktor berinteraksi dengan *use case.*

Selain Aktor, terdapat simbol-simbol lain dalam *Use Case* yang digunakan untuk memodelkan fungsi-fungsi sistem yang dibangun serta hubungan antar fungsinya. Beberapa simbol dalam *Use Case* antara lain sebagai berikut :

Tabel 2. 2 *Use Case Diagram*

Gambar	**Function**	**Keterangan**
	Asosiasi	digambarkan dengan garis yang menunjukkan siapa atau apa yang terlibat dalam interaksi secara langsung.

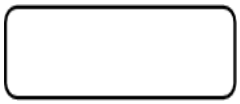
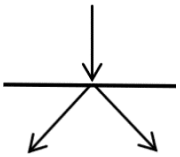
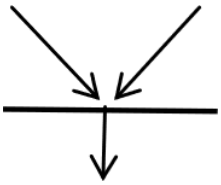
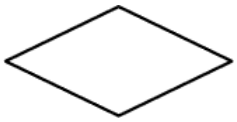
	Extend	Perluasan dari *use case* lain ketika beberapa kondisi dalam *use case* terpenuhi
	Include	Merupakan penambahan perilaku tambahan dalam suatu *use case*
	Use case generalization	Hubungan antara suatu use case umum dengan use case yang lebih spesifik.

2.7.3 Activity Diagram

Diagram aktivitas *(Activity Diagram)* menggambarkan aliran kerja dari sebuah sistem. Diagram aktivitas dapat mencakup cabang dan bercabangnya control dalam sebuah sistem yang berjalan secara bersamaan. Cabang cabang ini mewakili aktivitas yang dapat dilakukan secara bersamaan (Rumbaugh et al.., 2021). Simbol terkait adalah sebagai berikut :

Tabel 2. 3 Simbol *Activity Diagram*

Gambar	**Function**	**Keterangan**
	Start Point	Ditempatkan pada posisi sudut kiri atas yang menandakan awal mula dari aktifitas
	End Point	Penanda akhir dari suatu aktifitas

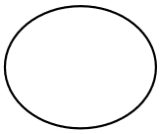
	*Activities*	Proses yang terjadi dalam aplikasi
	*Fork*	Menggambarkan kegiatan yang dilakukan secara bersamaan dalam aplikasi
	*Join*	Menandakan pemecahan aplikasi dalam bagian bagian yang lebih kecil
	*Decisions Point*	Menandakan pilihan untuk mengambil keputusajn jika dan maka

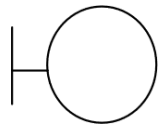
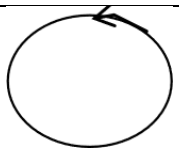
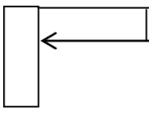
(Sumber : Rumbaugh et al., 2021**).**

2.7.4 Sequence Diagram

Sequence Diagram menjelaskan perilaku objek objek yang ada didalam *use case* dengan menjelaskan siklus hidup dari objek serta pesan pesan atau data yang dikirim oleh objek tersebut. Simbol terkait dalam *sequence diagram* meliputi :

Tabel 2. 4 Simbol *Sequence Diagram*

Simbol	**Function**	**Makna Simbol**
	*Entity Class*	Bagian dari sistem yang membentuk gambaran awal dari sistem

	*Boundary Class*	Berisi kumpulan kelas yang berinteraksi dengan satu atau lebih *actor* dalam sistem
	*Control Class*	Suatu entitas yang menampung logika pada aplikasi yang memiliki kewajiban langsung terhadap entitas lain
	Message	Simbol mengirim pesan antar class
	*Recursive*	Menggambarkan pengiriman pesan yang dikirim untuk dirinya sendiri

2.7.5 Class Diagram

Class diagram adalah salah satu jenis diagram struktur dalam UML yang dengan jelas menggambarkan struktur serta deskripsi kelas, atribut, metode, dan hubungan antara setiap objek. Class diagram bersifat statis, artinya diagram ini tidak menunjukkan apa yang terjadi saat kelas-kelas tersebut berinteraksi, melainkan menjelaskan jenis hubungan yang ada. *Class Diagram* disusun dalam komponen komponen yang dapat dilihat pada gambar dibawah ini.

Tabel 2. 5 Class Diagram

Simbol	Deskripsi	Simbol
Kelas (*Class*)	Menyatakan sebuah kelas dalam sistem. Kelas ini memuat atribut dan metode yang dimiliki oleh objek dari kelas tersebut.	Persegi panjang
Atribut (*Attribute*)	Menyatakan karakteristik atau properti dari objek yang diciptakan dari kelas tersebut.	Teks di dalam kelas
Metode (*Method/Operation*)	Menyatakan fungsi atau tindakan yang dapat dilakukan oleh objek kelas tersebut.	Teks di dalam kelas
Hubungan Asosiasi (*Association*)	Menunjukkan hubungan antara dua kelas. Terkadang dilengkapi dengan arah untuk menunjukkan arah hubungan.	Garis lurus
Hubungan Agregasi (*Aggregation*)	Menunjukkan hubungan **"bagian-dari"** antara dua kelas di mana objek dari satu kelas merupakan bagian dari objek kelas lainnya.	Garis dengan rombus di ujung
Hubungan Komposisi (*Composition*)	Merupakan bentuk agregasi yang lebih kuat, di mana objek bagian tidak dapat ada tanpa objek induknya.	Garis dengan rombus berwarna hitam di ujung
Hubungan Pewarisan (*Inheritance*)	Menunjukkan bahwa satu kelas (kelas anak) mewarisi atribut dan metode dari kelas lain (kelas induk).	Garis dengan panah segitiga kosong
Hubungan Implementasi (*Realization*)	Menunjukkan bahwa kelas tersebut mengimplementasikan antarmuka (interface).	Garis dengan panah segitiga berwarna penuh

Antarmuka (*Interface*)	Digunakan untuk mendefinisikan kontrak atau antarmuka yang akan diimplementasikan oleh kelas-kelas lainnya.	Persegi panjang dengan nama di bagian atas dan kata "interface"
Hubungan Ketergantungan (*Dependency*)	Menunjukkan bahwa suatu kelas bergantung pada kelas lain untuk menjalankan fungsinya.	Garis putus-putus dengan panah
Hubungan Multiplikasi (*Multiplicity*)	Menunjukkan jumlah objek yang dapat terhubung antara dua kelas.	Angka atau tanda bintang (*) pada garis asosiasi

2.8 Penelitian Terdahulu

Berbagai penelitian telah dilakukan dalam rangka mengembangkan sistem klasifikasi berbasis *algoritma C4.5* dan *Naïve Bayes*, khususnya dalam konteks pendidikan dan pengambilan Keputusan, dapat dilihat pada tabel dibawah ini.

Tabel 2. 6 Penelitian Terdahulu

No.	Penulis & Judul	Fokus Utama	Temuan Kunci	Relevansi & Celah Penelitian
1	(Rahmayanti et al., 2022) *Perbandingan Metode Algoritma C4.5 Dan Naïve Bayes Untuk*	Membandingkan akurasi C4.5 vs Naïve Bayes untuk prediksi kelulusan.	C4.5 (Akurasi 90%) lebih unggul dari Naïve Bayes (Akurasi 85%) pada dataset	Memberikan dasar perbandingan (baseline) kinerja kedua algoritma secara terpisah. Celah:

	Memprediksi Kelulusan Mahasiswa		kelulusan mahasiswa.	Penelitian ini hanya membandingkan, bukan mengintegrasikan kekuatan keduanya.
2	(Budiawan Zulfikar et al., n.d.) *Comparison of Naïve Bayes Classifier and C4.5 in Predicting Student Study Period*	Memprediksi masa studi mahasiswa menggunakan Naïve Bayes dan C4.5.	Naïve Bayes (Akurasi 96,12%) menunjukkan akurasi lebih tinggi dibanding C4.5 (93,82%).	Menunjukkan bahwa performa algoritma dapat bervariasi tergantung dataset. Ini memperkuat perlunya pendekatan yang lebih stabil seperti hybrid.
3	(Saputra et al., 2021) *A Comparative Analysis of C4.5, Naïve Bayes, and SVM...*	Membandingkan tiga algoritma untuk prediksi penyakit jantung.	Naïve Bayes yang dioptimalkan dengan PSO mencapai akurasi tertinggi (86,30%).	Mengindikasikan bahwa kinerja Naïve Bayes dapat ditingkatkan secara signifikan melalui pra-pemrosesan atau optimasi,

				sejalan dengan ide seleksi fitur.
4	(Changpetch et al., 2021) *Integrating Data Mining Techniques for Naïve Bayes Classification*	Meningkatkan Naïve Bayes dengan teknik pohon klasifikasi (mirip C4.5).	Integrasi pohon klasifikasi dengan Naïve Bayes terbukti mengunggul i teknik lain pada dataset medis.	Sangat Relevan. Memberikan bukti kuat bahwa menggunakan pohon keputusan sebagai pre-processor untuk Naïve Bayes adalah pendekatan yang efektif dan menjanjikan.

5	(Alharbi & Allohibi, 2024b) *A New Hybrid Classification Algorithm for Predicting Student Performance*	Mengembangkan algoritma hybrid baru untuk prediksi kinerja siswa.	Model hybrid baru (PHC) mengunggul i semua algoritma tunggal termasuk C4.5 dan Naïve Bayes.	Mendukung hipotesis umum bahwa model hybrid lebih superior daripada model tunggal dalam domain pendidikan.
6	(Rim & Liu, 2020) *Optimizing the C4.5 Decision Tree Algorithm Using MSD-Splitting*	Meningkatkan efisiensi dan akurasi C4.5 pada dataset besar.	Teknik optimasi MSD-Splitting berhasil meningkatk an akurasi dan mengurangi waktu eksekusi secara drastis.	Menegaskan kekuatan C4.5 dalam analisis dan seleksi fitur yang efisien, mendukung penggunaannya sebagai langkah awal dalam model hybrid.

7	(Chen et al., 2021) *Improved Naive Bayes Classification Algorithm*...	Meningkatkan Naïve Bayes dengan pembobotan fitur (feature weighting).	Akurasi meningkat signifikan (>99%) setelah fitur diberi bobot.	Menggarisbawahi kelemahan Naïve Bayes pada asumsi independensi fitur dan menunjukkan pentingnya seleksi atau pembobotan fitur sebelum klasifikasi.

BAB III
METOLODOGI PENELITIAN

3.1 Jenis Penelitian

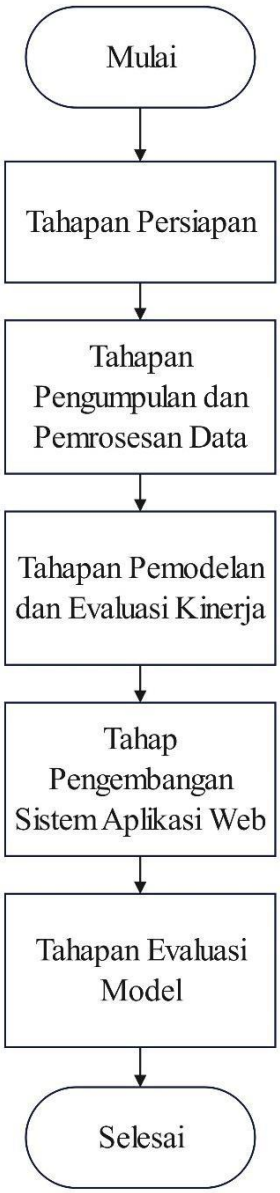
Penelitian ini menggunakan pendekatan kuantitatif. Pendekatan ini dipilih karena tujuan penelitian adalah untuk mengukur dan menganalisis kinerja dari model rekomendasi yang menerapkan algoritma C4.5 dan Naïve Bayes. Fokus utama dari pendekatan ini adalah pada analisis data numerik yang dihasilkan oleh model. Secara spesifik, penelitian akan mengukur metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai secara objektif seberapa efektif sistem dalam memberikan rekomendasi jurusan kepada siswa di SMA Negeri 8 Medan.

3.2 Lokasi Penelitian

Penelitian ini akan dilaksanakan di SMA Negeri 8 Medan, sebuah lokasi yang dipilih secara strategis karena adanya kebutuhan institusional untuk menyempurnakan sistem penjurusan melalui pendekatan berbasis data. Pemilihan ini juga didukung oleh ketersediaan data historis yang krusial dari sekolah, meliputi nilai akademik serta data minat studi siswa yang akan menjadi input utama bagi model. Lebih lanjut, jumlah siswa yang besar dan beragam di sekolah ini menyediakan sampel data yang dianggap representatif untuk melatih dan menguji model rekomendasi secara efektif.

3.3 Tahapan Penelitian

Untuk memastikan penelitian ini berjalan dengan baik, dibutuhkan beberapa langkah-langkah tahapan penelitian yang harus dilalui, secara garis besar, tahapana penelitian terkait dapat dilihat pada gambar dibawah ini.



Gambar 3. 1 *Flowchart* Penelitian

1. Tahapan Persiapan

Pada tahap awal, dilakukan studi literatur untuk mendalami teori mengenai algoritma C4.5, Naïve Bayes, dan sistem pendukung keputusan. Peneliti juga melakukan koordinasi dengan pihak SMA Negeri 8 Medan untuk perizinan dan identifikasi variabel data yang akan digunakan

2. Tahapan Pengumpulan dan Pra-pemrosesan data

Tahap ini berfokus pada pengumpulan data mentah dari sekolah, yaitu data nilai akademik dan hasil tes minat siswa. Setelah data terkumpul, dilakukan proses pra-pemrosesan yang mencakup pembersihan data dari nilai yang hilang atau tidak valid (*data cleaning*) dan standardisasi format data.

3. Tahapan Pemodelan dan Evaluasi Kinerja

Pada tahap pemodelan, fokus utamanya adalah membangun inti kecerdasan dari sistem rekomendasi. Proses ini diawali dengan seleksi fitur menggunakan algoritma C4.5, yang bekerja seperti seorang detektif untuk menyaring dan memilih atribut-atribut paling penting dari data siswa, misalnya nilai mata pelajaran kunci atau skor minat tertentu. Setelah hanya fitur-fitur yang paling berpengaruh yang tersisa, data tersebut kemudian digunakan untuk melatih model klasifikasi Naïve Bayes. Dalam pelatihan ini, model akan "belajar" dari sebagian besar data (data latih) untuk mengenali pola probabilitas antara profil siswa dengan jurusan yang sesuai. Setelah model selesai dilatih, tahap selanjutnya adalah evaluasi kinerja untuk mengukur seberapa baik model tersebut bekerja. Model akan diuji menggunakan sisa data yang belum pernah dilihat sebelumnya (data uji). Hasil prediksi dari model kemudian dibandingkan dengan hasil sebenarnya untuk menghitung metrik-metrik kuantitatif seperti accuracy (tingkat kebenaran secara keseluruhan), precision (ketepatan prediksi positif), dan recall (kemampuan model menemukan semua hasil positif). Semua perhitungan ini didasarkan pada confusion matrix, yang memberikan gambaran rinci tentang performa model dan menjadi dasar untuk menentukan apakah model tersebut cukup andal untuk digunakan

4. Tahap Pengembangan Sistem Berbasis Web

Tahap pengembangan sistem berbasis web adalah proses membangun *prototype* fungsional dengan mengintegrasikan model prediksi yang telah

divalidasi. Proses ini melibatkan pembangunan *backend* (logika sistem) menggunakan *Python* yang bertugas menyediakan API, memuat model .pkl yang telah disimpan untuk melakukan prediksi, dan mengelola seluruh interaksi dengan basis data MySQL.

Di sisi lain, *frontend* (antarmuka pengguna) akan dibangun menggunakan PHP untuk menyajikan halaman interaktif kepada pengguna. Kedua komponen ini akan bekerja secara terintegrasi. frontend mengirimkan data masukan dari pengguna ke API *backend*, yang kemudian memprosesnya dengan model dan mengirimkan kembali hasil rekomendasi untuk ditampilkan secara visual kepada pengguna.

5. Tahapan Evaluasi Model

Proses evaluasi ini dilakukan setelah model selesai dilatih dan sebelum diintegrasikan ke dalam aplikasi web. Berikut adalah langkah-langkah rincinya :

a) Pengujian dengan Data Uji : Model akan diberikan data uji (*testing set*), yaitu sebagian kecil dari data siswa yang sengaja tidak digunakan selama proses pelatihan. Model kemudian diminta untuk memprediksi jurusan untuk setiap siswa dalam data uji tersebut.
b) Penggunaan *Confussion Matrix* : Hasil prediksi dari model akan dibandingkan dengan jurusan yang sebenarnya dari data uji. Perbandingan ini diringkas dalam sebuah tabel yang disebut *confusion matrix*, yang secara detail menunjukkan empat kemungkinan hasil: prediksi yang benar (*True Positive* & *True Negative*) dan prediksi yang salah (*False Positive* & *False Negative*).
c) Perhitungan Metrik Kinerja : Berdasarkan angka-angka dari *confusion matrix*, Anda akan menghitung beberapa metrik statistik untuk menilai performa model dari berbagai sudut :
   (1) ***Accuracy*** (Akurasi) bermanfaat untuk Mengukur persentase prediksi yang benar secara keseluruhan. Ini menjawab pertanyaan, "Seberapa sering model menebak dengan benar?"

(2) ***Precision* (Presisi):** Mengukur tingkat ketepatan dari prediksi positif. Ini menjawab, "Dari semua siswa yang direkomendasikan masuk jurusan X, berapa persen yang benar-benar cocok?”

(3) ***Recall* (Daya Ingat):** Mengukur kemampuan model untuk menemukan semua kasus positif yang sebenarnya. Ini menjawab, "Dari semua siswa yang seharusnya masuk jurusan X, berapa persen yang berhasil ditemukan oleh model?"

(4) ***F1-Score*:** Merupakan nilai rata-rata yang menyeimbangkan antara *precision* dan *recall*, memberikan satu angka tunggal yang merepresentasikan kualitas model secara keseluruhan.

3.4 Data dan Sumber Data

Data yang digunakan dalam penelitian ini terdiri dari dua jenis data utama yang sangat penting untuk membangun sistem rekomendasi jurusan kuliah berbasis algoritma C4.5 dan Naïve Bayes. Jenis data yang digunakan adalah:

1. Data Nilai Akademik Siswa.

   Data ini mencakup nilai rapor siswa pada berbagai mata pelajaran penting, seperti Matematika, Bahasa Indonesia, Bahasa Inggris, dan mata pelajaran lainnya. Nilai-nilai ini akan digunakan untuk mengetahui kemampuan akademik siswa yang kemudian akan digunakan sebagai salah satu input untuk proses klasifikasi dalam sistem rekomendasi.

2. Data Hasil Tes Minat Siswa

   Data ini mencakup hasil tes minat yang dilakukan oleh pihak sekolah untuk mengetahui minat siswa terhadap jurusan kuliah tertentu. Tes minat ini merupakan salah satu variabel yang penting karena akan mempengaruhi rekomendasi jurusan yang diberikan kepada siswa. Data hasil tes ini akan digunakan sebagai salah satu fitur utama dalam sistem rekomendasi.

3.5 Teknik Pengumpulan Data

Tahapan pengumpulan data dalam penelitian ini dapat dilihat pada gambar dibawah ini.

Tabel 3. 1 Tahapan Pengumpulan Data

Tahap	Kegiatan	Instrumen	Hasil yang Diharapkan
1. Izin & Persiapan	Minta persetujuan kepala sekolah, Wali Kelas, serta orang tua. Susun jadwal pengambilan data.	Surat izin & lembar persetujuan	Dokumen izin lengkap dan jadwal jelas.
2. Pengumpulan Nilai Rapor	Ambil file **Excel** yang sudah berisi nilai rapor semester 1-5 dari TU.	File Excel nilai rapor	Data nilai rapor digital siap olah.
3. Tes Minat	Gunakan hasil tes minat yang **sudah diarsip** oleh BK. (Jika ada)	Rekap skor tes minat (Excel/PDF)	Skor minat belajar tiap siswa.

3.6 Teknik Pengolahan dan Analisis Data

Pengolahan dan analisis data dilakukan menggunakan bahasa *Python* dengan bantuan beberapa library seperti *Pandas* dan *Scikit-learn*. Berikut ini adalah langkah-langkah yang dilakukan sampai tahap pembagian data:

1. Menggabungkan Data

   Langkah awal dalam proses pengolahan data adalah menggabungkan semua sumber data utama yang diperoleh dari sekolah. Setelah digabung, semua informasi penting dari kedua sumber data akan tersedia dalam satu tabel utama (dataset). Tabel ini akan digunakan sebagai *input* untuk seluruh analisis dan pelatihan model, *library* yang digunakan adalah *pandas*.

2. Pembersihan dan Standardisasi Data

Setelah data digabung, langkah selanjutnya adalah membersihkannya dari nilai yang kosong atau tidak valid dan menyeragamkan skala penilaian. Baris yang memiliki nilai kosong pada kolom penting seperti Matematika, IPA, atau IPS akan dihapus agar tidak mengganggu proses pelatihan model. Selain itu, jika ada kolom nilai yang menggunakan skala berbeda (misalnya skala 1–4), nilainya akan dikonversi ke skala 0–100 untuk menyamakan bobot antar mata pelajaran.

Proses ini dilakukan menggunakan library **pandas**, yang memungkinkan peneliti memeriksa, memfilter, dan mengubah isi tabel dengan mudah. Fungsi seperti dropna() digunakan untuk menghapus baris yang tidak lengkap, sedangkan apply() atau operasi vektor digunakan untuk menyesuaikan skala nilai secara efisien di seluruh kolom. Dengan data yang bersih dan seragam, model yang dilatih nantinya akan menghasilkan prediksi yang lebih akurat dan tidak bias.

3. Seleksi Fitur dengan *Decision Tree (C4.5)*

Langkah ini bertujuan untuk memilih fitur-fitur (kolom) yang paling berpengaruh dalam menentukan jurusan siswa, seperti nilai Matematika, IPA, atau skor minat tertentu. Seleksi fitur dilakukan menggunakan algoritma Decision Tree dengan kriteria entropy, yang mewakili prinsip dari metode C4.5. Model akan menghitung *feature importance* dari setiap fitur, lalu fitur-fitur yang memiliki kontribusi penting akan dipilih untuk digunakan dalam pelatihan model klasifikasi. Proses ini dilakukan menggunakan **DecisionTreeClassifier** dari library **scikit-learn**. Dengan pendekatan ini, peneliti dapat secara objektif menentukan fitur mana saja yang benar-benar relevan, sehingga proses pelatihan menjadi lebih efisien dan akurat, serta dapat mengurangi noise dari fitur yang tidak terlalu berpengaruh.

4. Pembagian Data Latih dan Uji (Train–Test Split)

Setelah fitur-fitur penting dipilih, data kemudian dibagi menjadi dua bagian: data latih dan data uji. Umumnya, pembagian dilakukan dengan proporsi 80% untuk data latih dan 20% untuk data uji. Data latih digunakan untuk membangun dan melatih model, sementara data uji digunakan untuk mengevaluasi seberapa baik model tersebut bekerja pada data yang belum pernah dilihat sebelumnya.

Proses ini dilakukan menggunakan fungsi train_test_split() dari library **scikit- learn**. Untuk menjaga keseimbangan jumlah siswa di setiap jurusan (IPA dan IPS), pembagian dilakukan dengan parameter **stratify**, sehingga distribusi kelas tetap proporsional di kedua subset data. Pembagian yang tepat ini penting agar evaluasi model menjadi lebih adil dan akurat.

5. Pelatihan Model Naïve Bayes

Setelah data dibagi, model klasifikasi dilatih menggunakan algoritma *Naïve Bayes*, tepatnya versi *Gaussian Naïve Bayes* yang cocok untuk data numerik seperti nilai rapor dan skor tes minat. Model ini bekerja berdasarkan prinsip probabilistik, menghitung peluang masing-masing kelas berdasarkan distribusi nilai dari fitur- fitur yang telah dipilih sebelumnya.

Proses pelatihan dilakukan menggunakan kelas GaussianNB dari library **scikit- learn**. Model dilatih dengan data latih, kemudian diuji menggunakan data uji untuk memprediksi jurusan siswa. Naïve Bayes dipilih karena sederhana, cepat, dan cukup efektif dalam menangani masalah klasifikasi dengan jumlah data terbatas seperti dalam penelitian ini.

6. Evaluasi Model

Setelah model Naïve Bayes dilatih, langkah selanjutnya adalah mengevaluasi kinerjanya untuk melihat seberapa baik model dalam memprediksi jurusan siswa. Evaluasi dilakukan dengan membandingkan hasil prediksi pada data uji dengan label asli jurusan menggunakan metrik

seperti akurasi, *precision*, *recall*, dan *F1-score*. Selain itu, ***confusion matrix*** juga digunakan untuk mengetahui jumlah prediksi yang benar dan salah secara rinci.

Evaluasi ini dilakukan menggunakan fungsi-fungsi dari modul **sklearn.metrics**, seperti accuracy_score(), classification_report(), dan confusion_matrix(). Hasil evaluasi ini akan menjadi dasar untuk menilai apakah model sudah cukup baik digunakan dalam sistem rekomendasi atau perlu dilakukan perbaikan lebih lanjut, misalnya dengan menambah data atau mengubah metode klasifikasi.

7. Penyimpanan Model

Setelah model berhasil dilatih dan dievaluasi, langkah terakhir adalah menyimpan model agar dapat digunakan kembali tanpa perlu melatih ulang. Model disimpan dalam format file .pkl (pickle), yang memungkinkan sistem atau aplikasi rekomendasi jurusan memuat dan menggunakan model secara langsung untuk memproses data siswa baru.

Proses penyimpanan ini dilakukan menggunakan library **joblib**, yang dirancang khusus untuk menyimpan objek berukuran besar seperti model machine learning. Selain model itu sendiri, daftar fitur yang digunakan juga disimpan agar sistem nantinya hanya menggunakan input yang sesuai. Dengan demikian, model dapat dengan mudah diintegrasikan ke dalam antarmuka web atau aplikasi berbasis Python secara efisien dan konsisten.

8. Implementasi Model dalam Sistem

Setelah model disimpan, langkah berikutnya adalah mengintegrasikan model ke dalam sistem rekomendasi jurusan. Model yang telah disimpan akan dimuat ulang (load) setiap kali sistem menerima input baru dari siswa, seperti nilai rapor dan hasil tes minat. Data tersebut akan diproses menggunakan fitur yang telah dipilih sebelumnya, kemudian diprediksi jurusannya menggunakan model Naïve Bayes.

Library **joblib** kembali digunakan untuk memuat model (load()), sedangkan **pandas** digunakan untuk menyiapkan data input dalam format DataFrame sesuai struktur pelatihan. Proses ini membuat sistem dapat memberikan hasil rekomendasi secara otomatis dan real-time, tanpa perlu melatih model ulang setiap kali ada data baru. Hasil rekomendasi dapat ditampilkan langsung kepada siswa atau Wali Kelas dalam bentuk visual atau laporan tertulis.

3.7 Perancangan Sistem

Perancangan sistem dalam penelitian ini bertujuan untuk membangun aplikasi rekomendasi jurusan kuliah berbasis algoritma *C4.5* dan *Naïve Bayes*, yang dapat membantu siswa dan Wali Kelas dalam menentukan jurusan kuliah yang sesuai berdasarkan data nilai akademik dan hasil tes minat siswa.

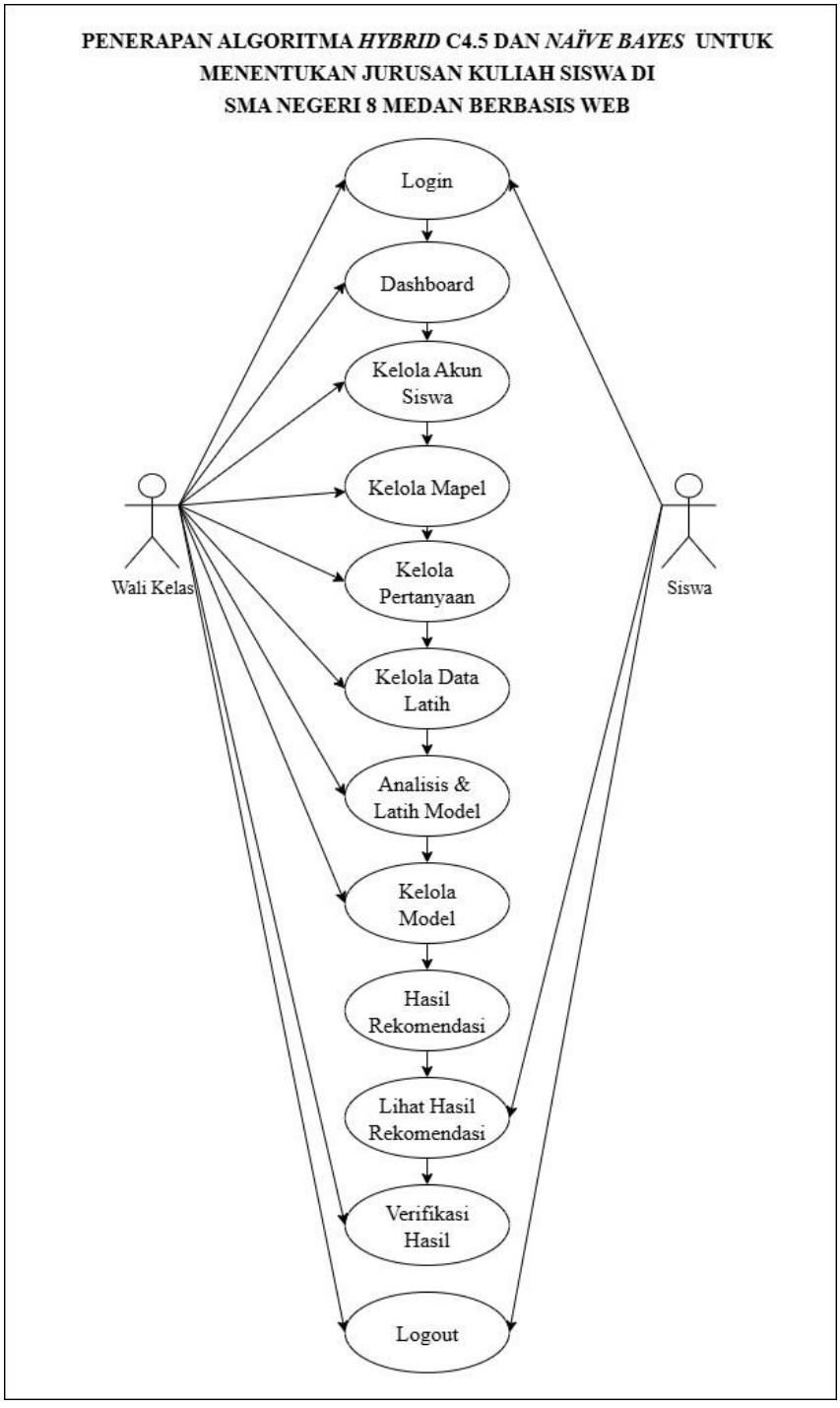
Dalam merancang sistem ini, digunakan pendekatan *Unified Modeling Language* (UML) untuk menggambarkan struktur dan alur sistem secara terperinci. UML merupakan standar internasional yang digunakan dalam analisis dan desain sistem berorientasi objek. Dengan UML, sistem dapat didokumentasikan secara visual agar lebih mudah dipahami oleh pengembang maupun pihak sekolah yang akan menggunakan aplikasi ini.

Adapun jenis diagram UML yang digunakan dalam perancangan sistem ini meliputi :

1. *Use Case Diagram* - Menggambarkan fungsi-fungsi utama yang dapat dilakukan oleh pengguna (aktor) terhadap sistem.
2. *Activity Diagram* – Menunjukkan alur aktivitas dalam sistem, seperti proses pengisian data siswa hingga pemberian rekomendasi jurusan
3. *Class Diagram* – Merepresentasikan struktur data dalam sistem, termasuk atribut dan relasi antar entitat seperti siswa, nilai, hasil minat, dan rekomendasi.
4. *Sequence Diagram* – Menggambarkan urutan interaksi antar objek dalam sistem saat proses rekomendasi dijalankan.

3.7.1. *Use Case Diagram*

Perancangan *Use Case Diagram* dapat dilihat pada gambar dibawah ini



Gambar 3. 2 *Use Case Diagram*

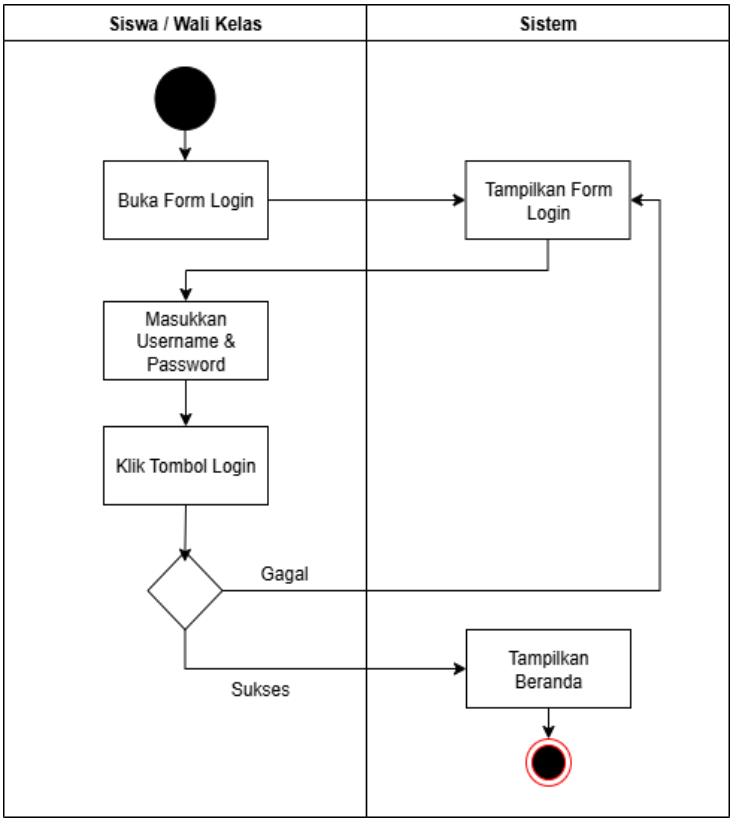
Dalam *Use case diagram* dibawah ini, sistem ini melibatkan

interaksi antara dua aktor utama: **Wali Kelas** dan **Siswa**. Wali Kelas memiliki hak akses penuh untuk mengelola seluruh fungsionalitas sistem, mulai dari login, mengelola data nilai dan tes minat siswa, menjalankan proses pengolahan data dengan algoritma, melihat seluruh hasil rekomendasi, melakukan verifikasi, hingga manajemen data akun siswa. Di sisi lain, aktor Siswa memiliki akses yang lebih terbatas, yaitu hanya untuk login, melihat hasil dan riwayat rekomendasi yang ditujukan secara spesifik untuk dirinya, dan kemudian logout dari sistem.

3.7.2. *Activity Diagram*

1. *Activity Diagram Login*

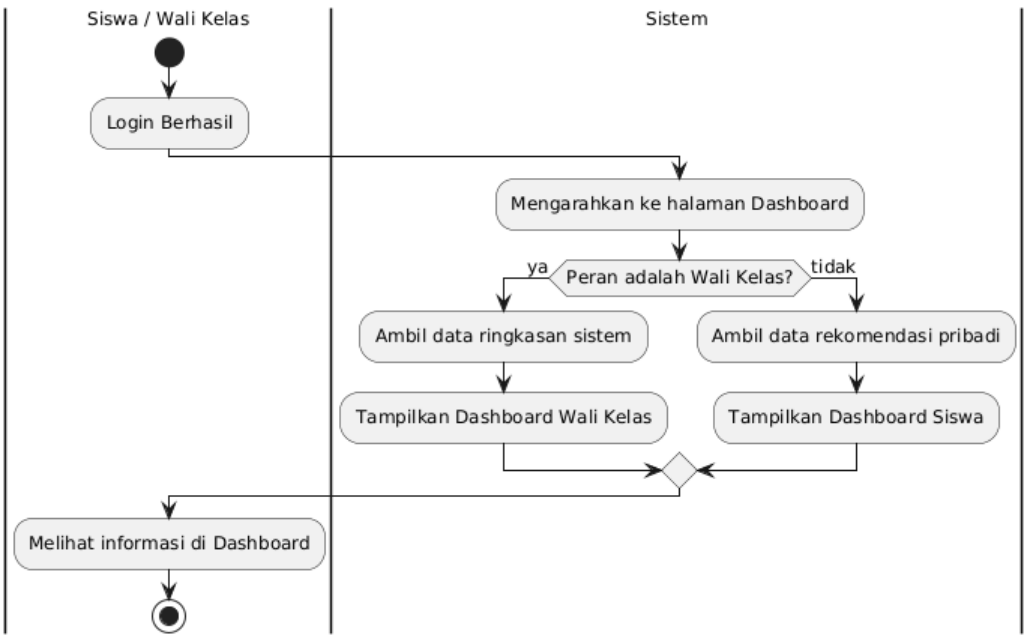
   Diagram ini menunjukkan proses masuk untuk pengguna, yaitu Siswa atau Wali Kelas. Pengguna memasukkan kredensial yang kemudian divalidasi oleh sistem. Jika berhasil, sistem akan mengarahkan ke dasbor yang sesuai; jika gagal, pesan kesalahan akan ditampilkan agar pengguna dapat mencoba lagi.



Gambar 3. 3 *Activity Diagram Login*

2. *Activity Diagram* Dashboard

Diagram ini menggambarkan aktivitas setelah pengguna berhasil login. Sistem akan menampilkan dashboard yang kontennya disesuaikan berdasarkan peran pengguna, apakah sebagai Wali Kelas atau Siswa.



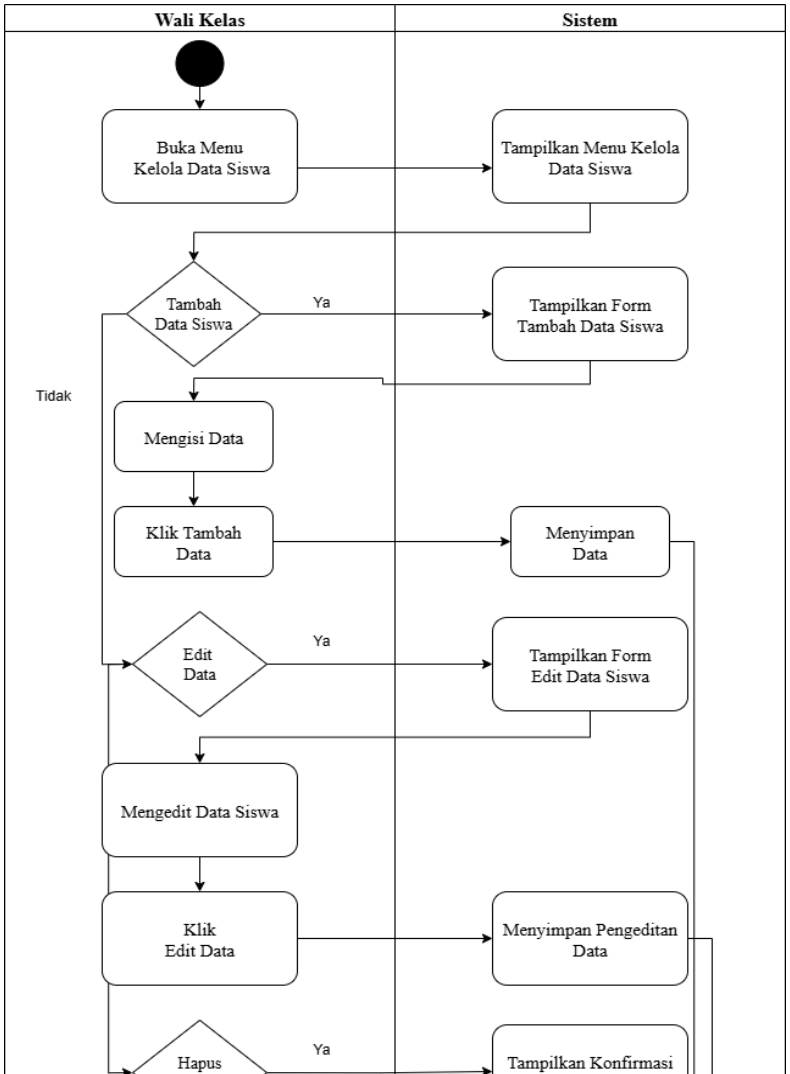
Gambar 3. 4 Activity Diagram *Dashboard*

3. *Activity Diagram Kelola* Akun Siswa

Diagram ini merinci alur kerja Wali Kelas dalam mengelola data akun siswa. Aktivitas yang dicakup meliputi proses menambah siswa baru, mengubah data siswa yang ada, dan menghapus akun siswa dari sistem.

4. *Activity Diagram* Kelola Akun Siswa

Diagram ini merinci alur kerja Wali Kelas dalam mengelola data akun siswa. Aktivitas yang dicakup meliputi proses menambah siswa baru, mengubah data siswa yang ada, dan menghapus akun siswa dari sistem.

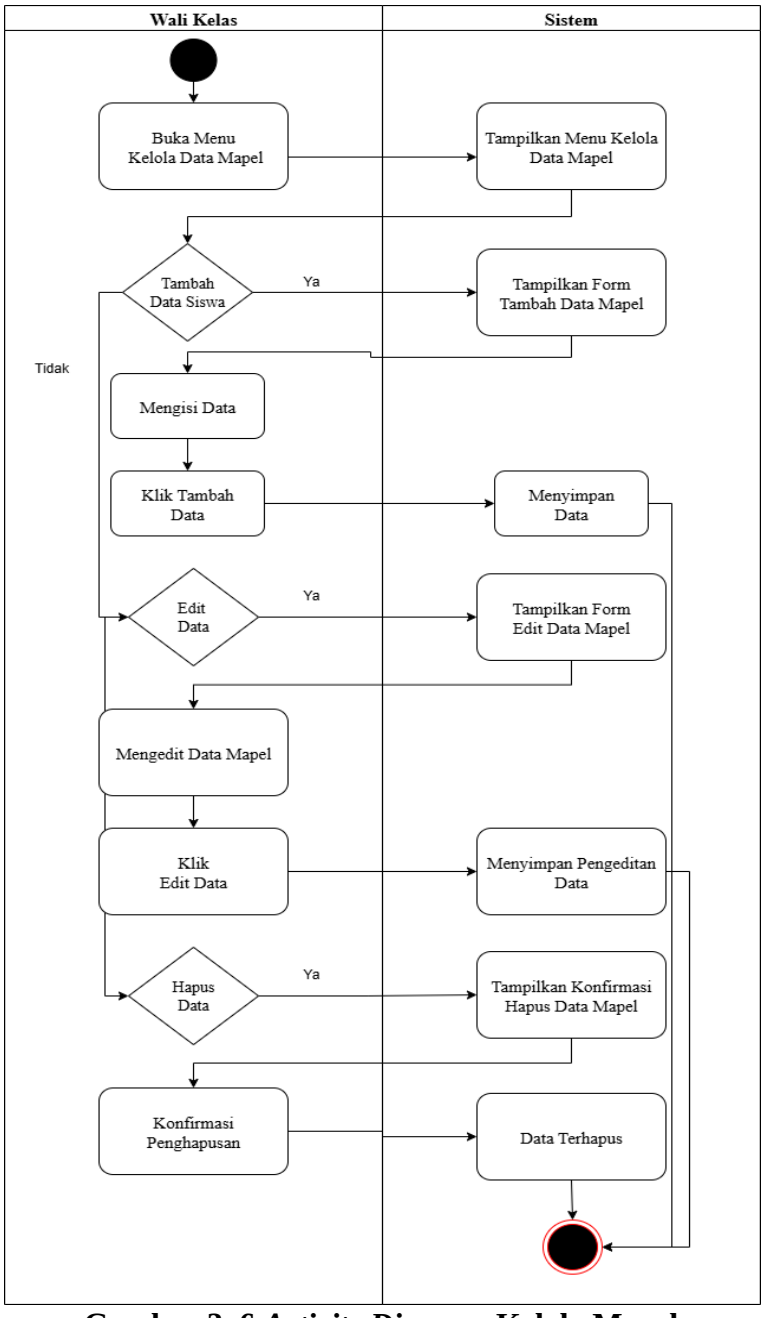


Gambar 3. 5 Activity Diagram Kelola Akun Siswa

Gambar 3. 5 Activity Diagram Kelola Akun Siswa

5. *Activity* Diagram Kelola Mapel

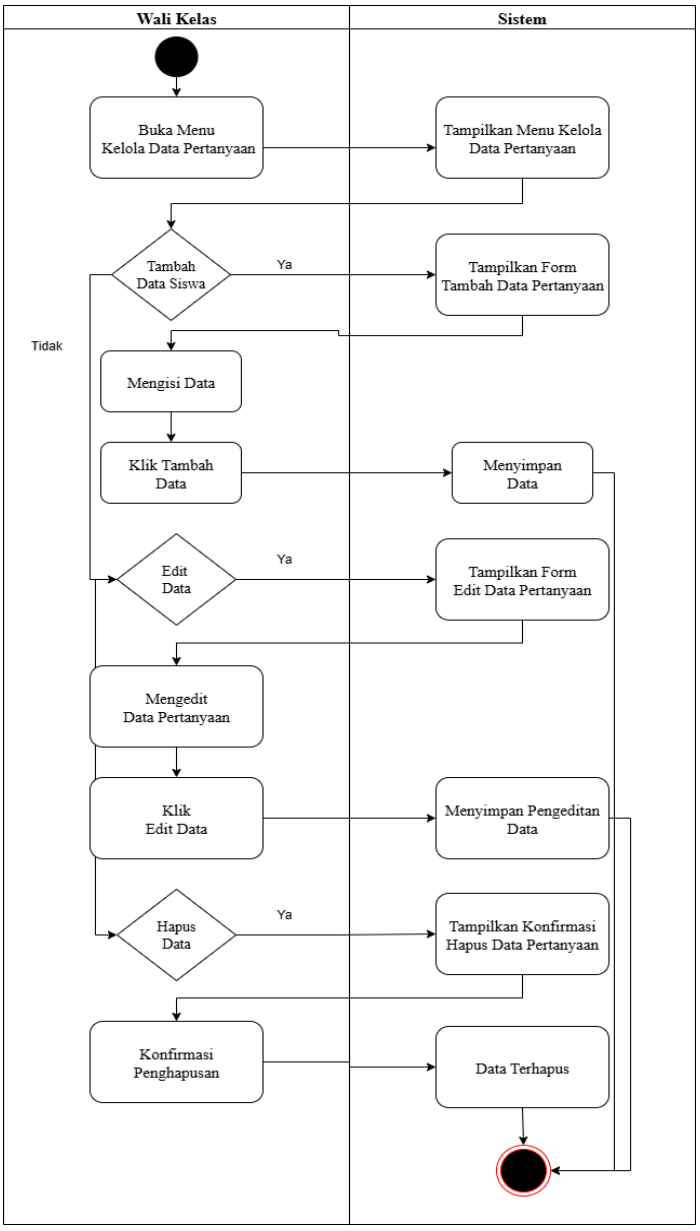
Diagram ini menjelaskan alur pengelolaan data latih oleh Wali Kelas, yang mencakup data nilai (mapel). Proses ini meliputi penambahan, pengubahan, dan penghapusan data yang menjadi dasar untuk analisis model.



Gambar 3. 6 *Activity Diagram* Kelola Mapel

6. *Activity Diagram* Kelola Pertanyaan

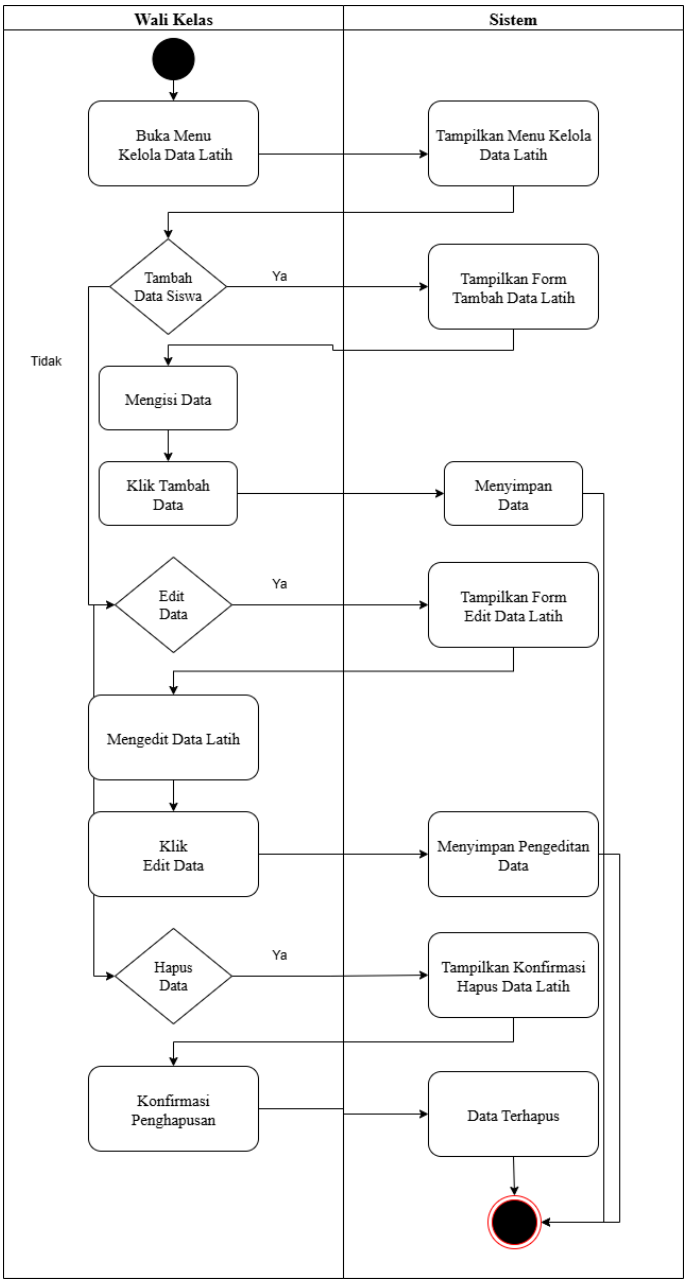
Diagram ini menjelaskan alur pengelolaan data pertanyaan oleh Wali Kelas, yang mencakup data pertanyaan minat. Proses ini meliputi penambahan, pengubahan, dan penghapusan data yang menjadi dasar untuk analisis model.



Gambar 3. 7 Activity Diagram Kelola Pertanyaan

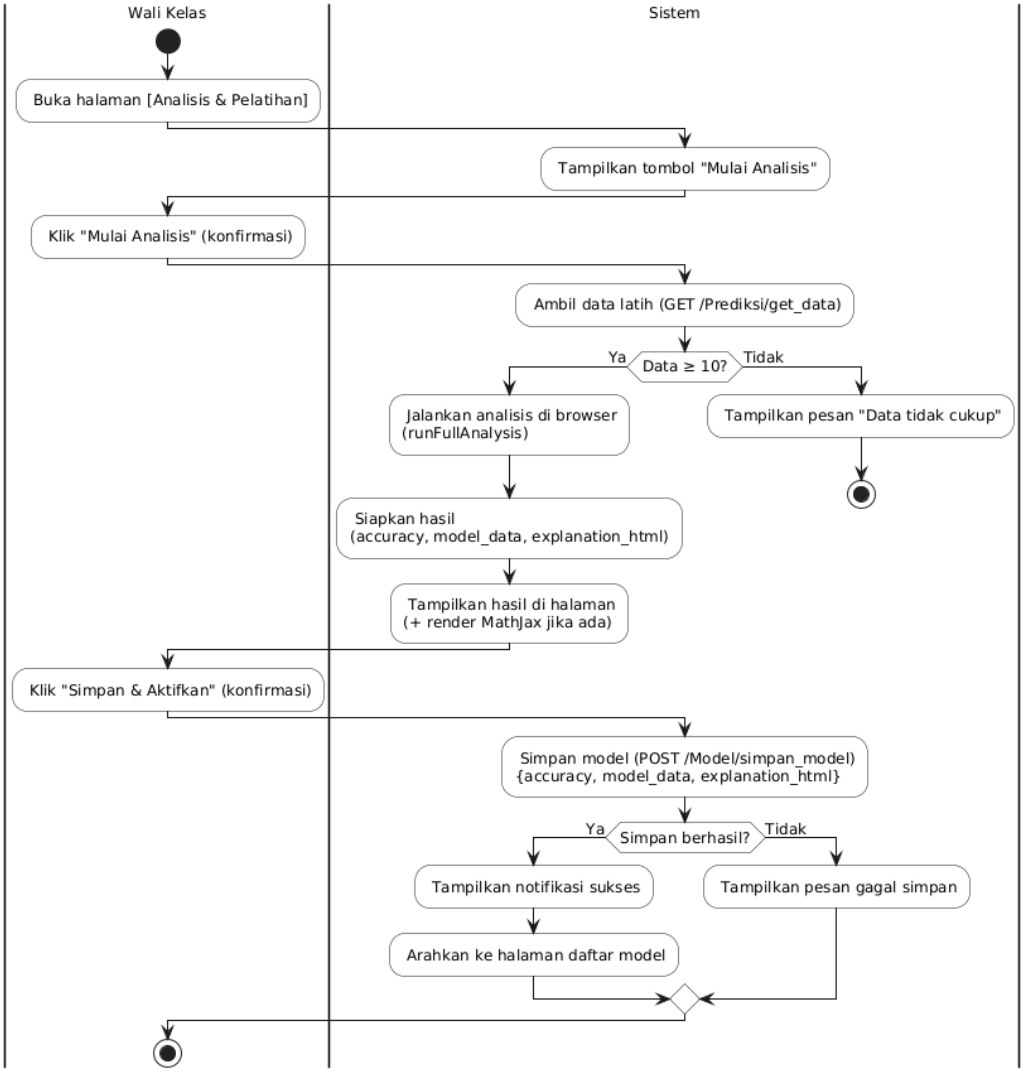
7. *Activity Diagram* Kelola Data Latih

Diagram ini menjelaskan alur pengelolaan data latih oleh Wali Kelas, yang mencakup data pertanyaan minat. Proses ini meliputi penambahan, pengubahan, dan penghapusan data yang menjadi dasar untuk analisis model.



Gambar 3. 8 *Activity Diagram* Kelola Data Latih

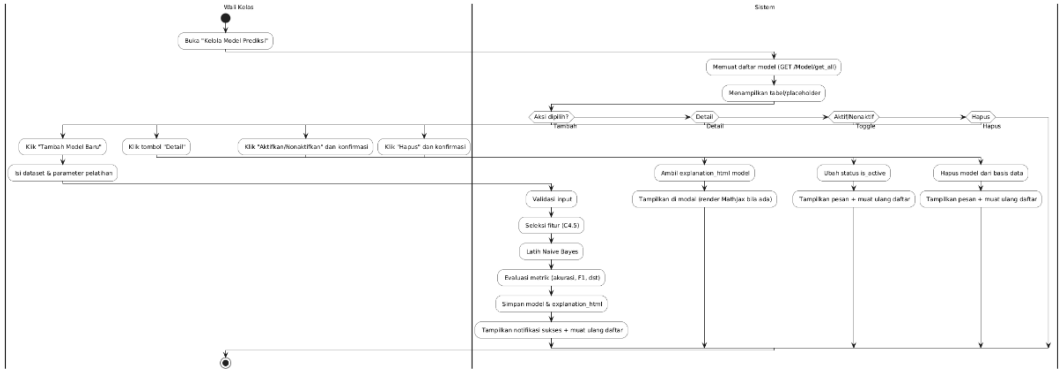
8. *Activity Diagram* Analisis dan Latih Model



Gambar 3. 9 *Activity Diagram* Analisis dan Latih Model

9. *Activity Diagram* Kelola Data Model

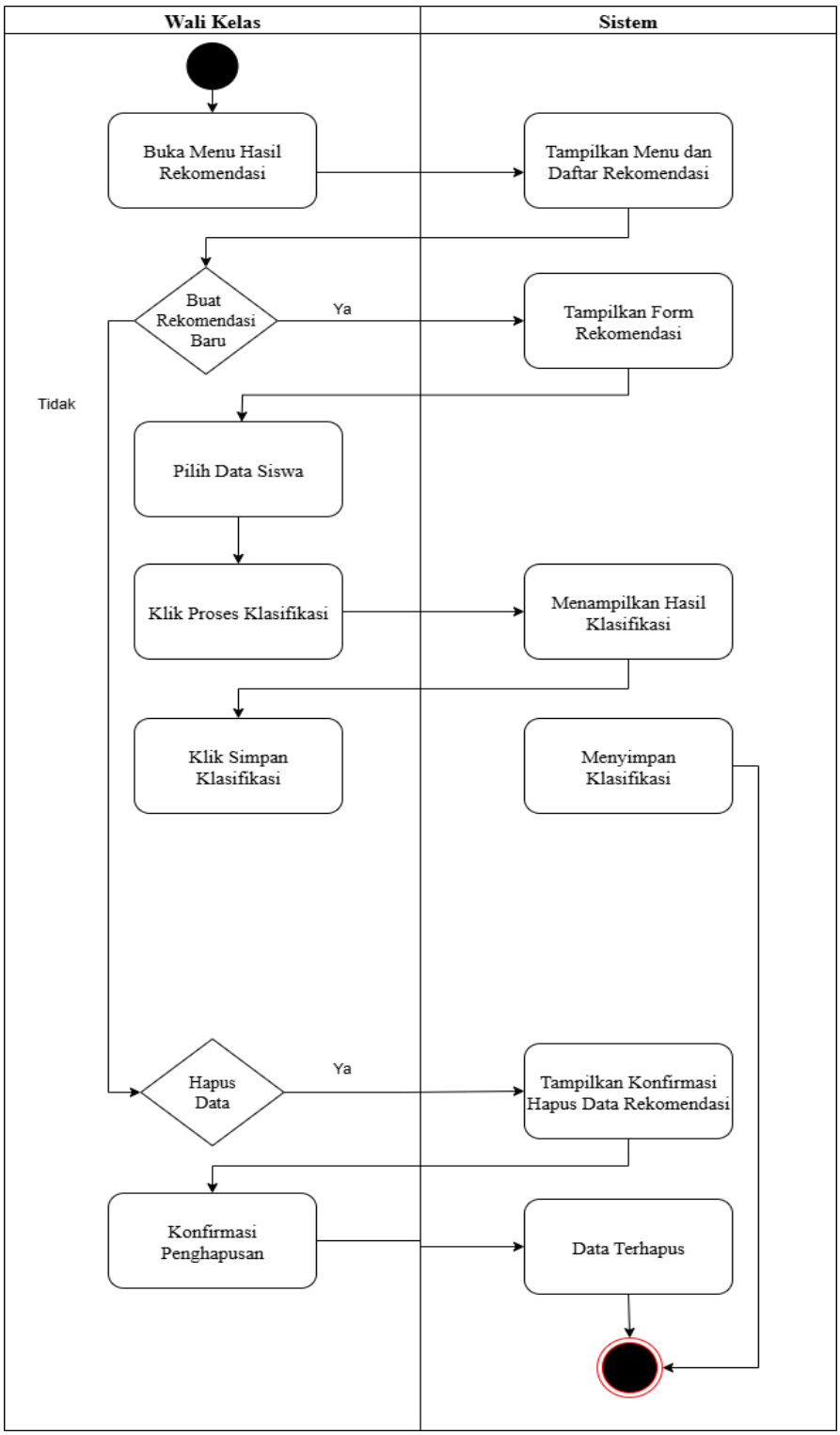
Diagram ini menyederhanakan alur jadi dua pihak saja: Wali Kelas dan Sistem. Wali Kelas memilih aksi (tambah/lihat detail/aktifkan-nonaktifkan/hapus), lalu Sistem memproses permintaan, menampilkan hasil, dan memuat ulang daftar model.



Gambar 3. 10 *Activity Diagram* Kelola Data Model

10. *Activity Diagram* Hasil Rekomendasi (Wali Kelas)

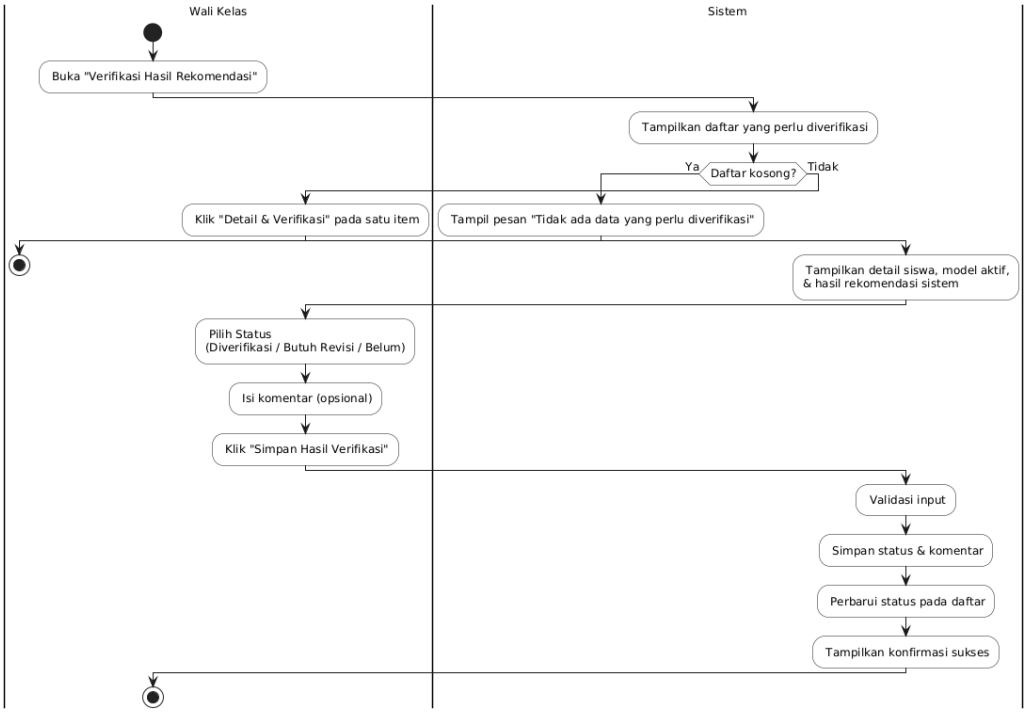
Diagram ini menggambarkan Wali Kelas membuka Hasil Rekomendasi, membuat rekomendasi baru dengan memilih data siswa, menjalankan klasifikasi di sistem, lalu menyimpan hasilnya. Diagram juga mencakup penghapusan rekomendasi yang diawali konfirmasi dari sistem sebelum data dihapus.



Gambar 3. 11 Activity Diagram Hasil Rekomendasi

11. *Activity Diagram* Verifikasi Hasil

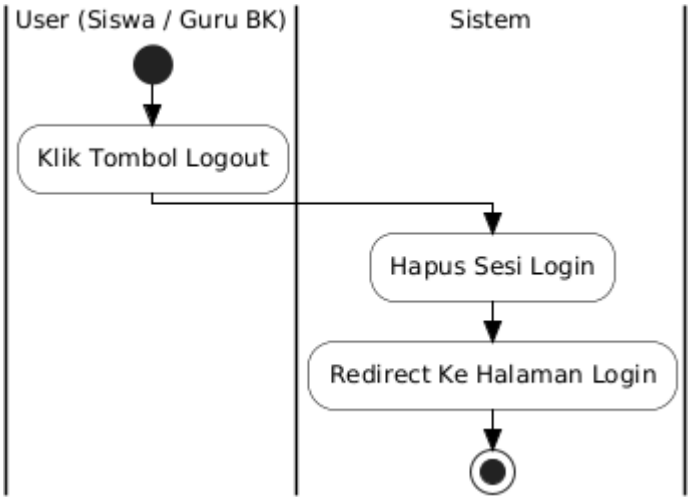
Dalam diagram ini, Wali Kelas membuka daftar verifikasi, memilih satu hasil, meninjau detail siswa serta rekomendasi model, lalu menentukan status (Diverifikasi/Butuh Revisi/Belum) dan menulis komentar. Sistem menyimpan keputusan dan catatan tersebut, memperbarui status pada daftar verifikasi, dan menampilkan konfirmasi.



Gambar 3. 12 Activity Diagram Verifikasi Hasil

12. *Activity Diagram Logout*

Diagram terakhir ini menunjukkan proses keluar dari sistem untuk Siswa atau Wali Kelas. Saat pengguna mengklik tombol logout, sistem akan mengakhiri sesi aktif secara aman. Setelah itu, pengguna akan dialihkan kembali ke halaman login utama.

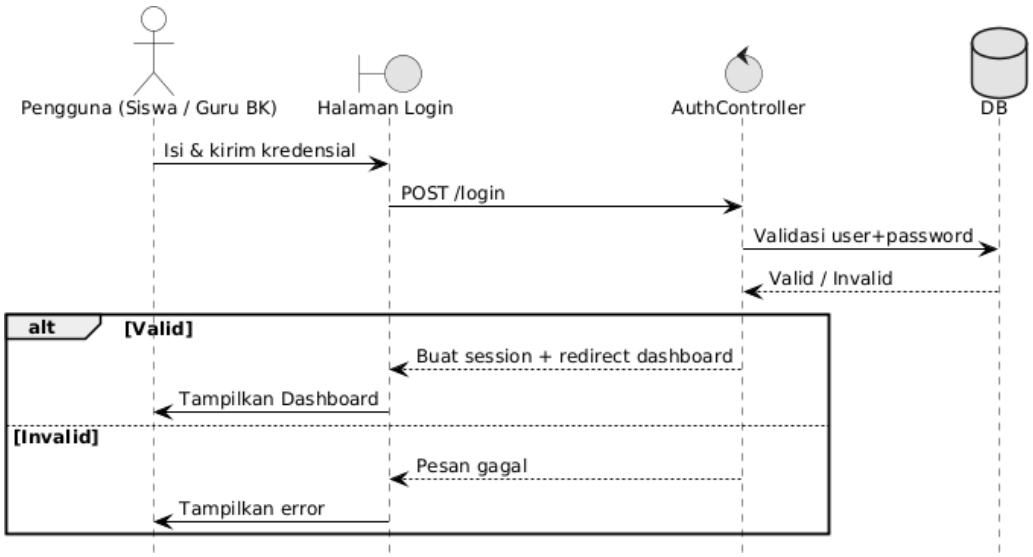


Gambar 3. 13 ***Activity Diagram Logout***

3.7.3. *Sequence Diagram*

1. *Sequence Diagram Login*

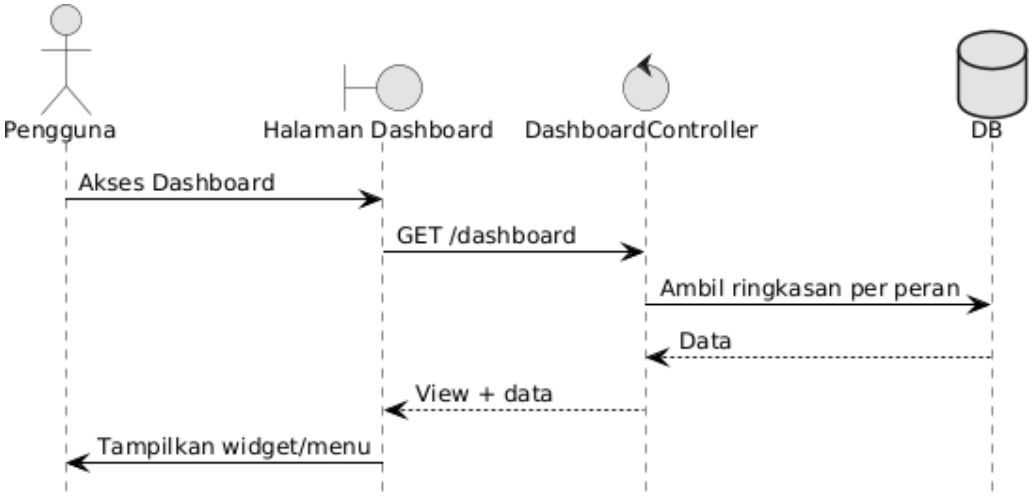
Diagram ini mengilustrasikan alur login. Pengguna mengirim kredensial; sistem memvalidasi ke DB lalu mengarahkan ke dashboard jika valid, atau menampilkan pesan gagal.



Gambar 3. 14 ***Sequence Diagram Login***

2. *Sequence Diagram Dashboard*

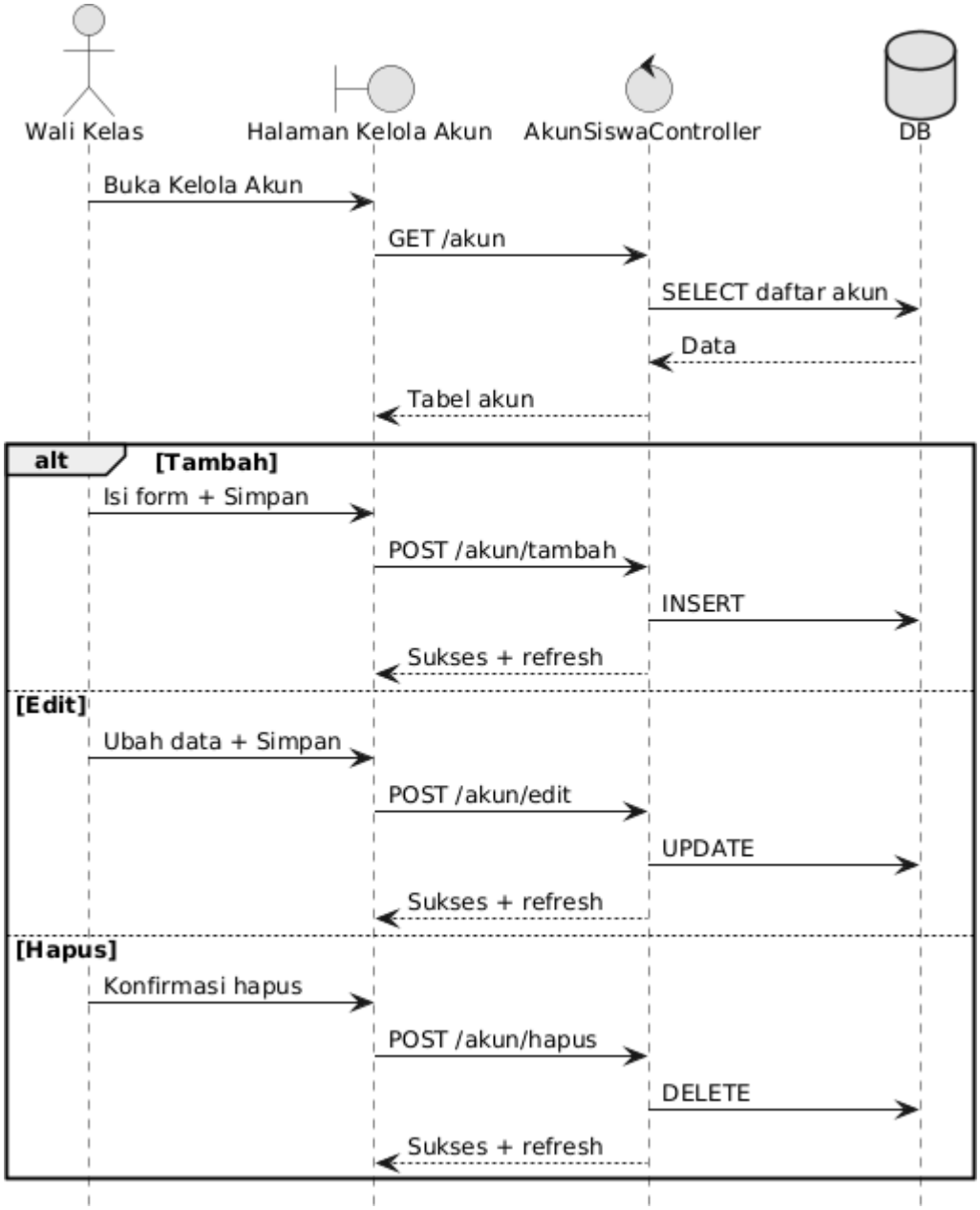
Setelah berhasil login, diagram ini menunjukkan bagaimana sistem menampilkan dasbor yang berbeda untuk **Wali Kelas** dan **Siswa**. Sistem akan mengambil data yang relevan dari database sesuai peran pengguna. Wali Kelas melihat ringkasan data siswa, sementara Siswa melihat hasil rekomendasi jurusannya.



Gambar 3. 15 *Sequence Diagram Dashboard*

3. *Sequence Diagram* Kelola Akun Kelas

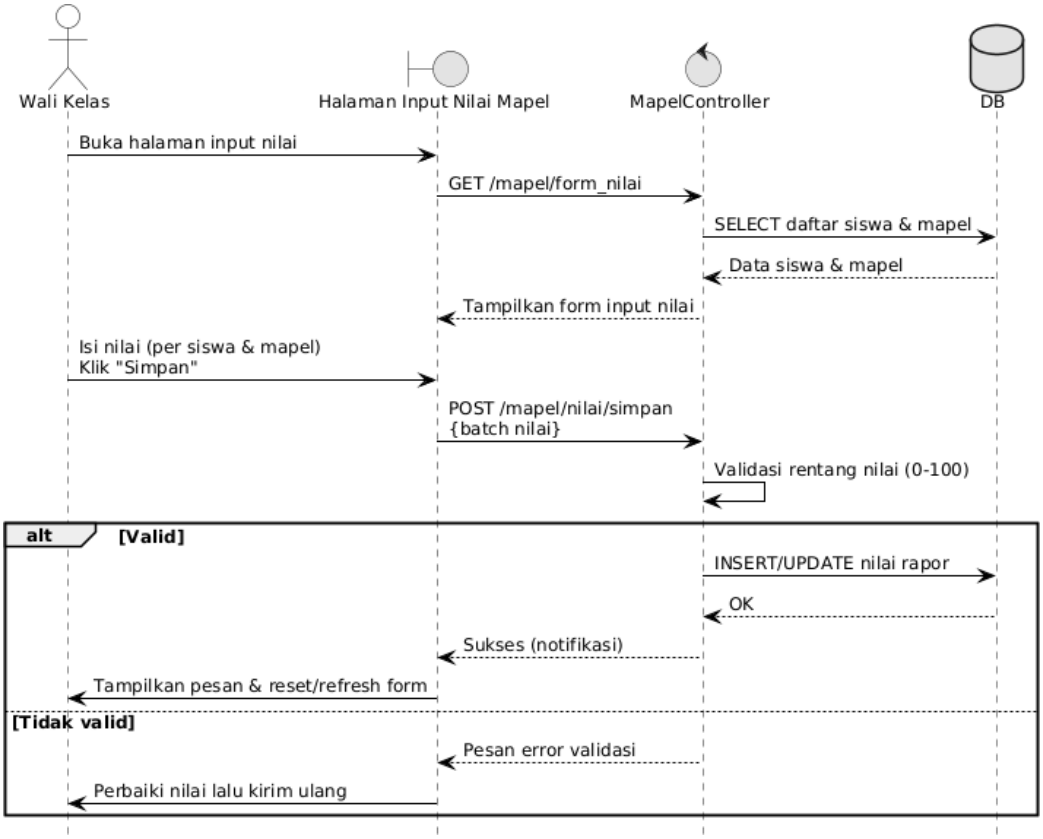
Dalam diagram dibawah ini, Wali Kelas melihat daftar akun lalu menambah, mengedit, atau menghapus; sistem menyimpan perubahan ke DB.



Gambar 3. 16 *Sequence Diagram* Kelola Akun Siswa

4. *Sequence Diagram* Kelola Mata Pelajaran

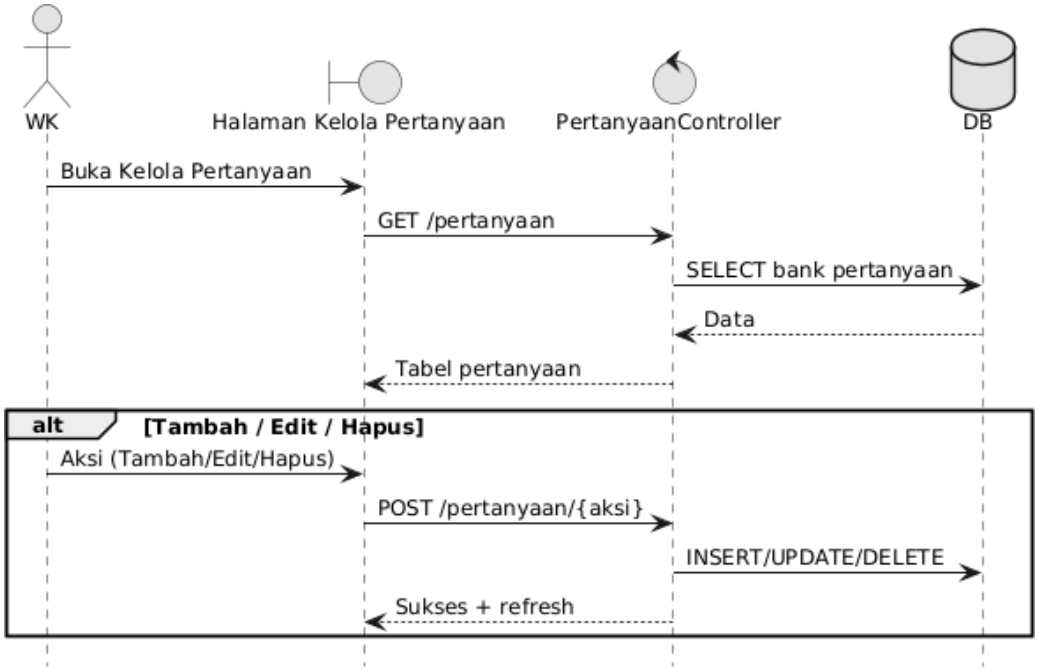
Diagram ini menjelaskan alur penambahan mata Pelajaran. Wali Kelas membuka form, mengisi nilai, lalu menyimpan. Sistem memuat daftar siswa/mapel, memvalidasi, menyimpan ke database, dan menampilkan konfirmasi.



Gambar 3. 17 Sequence Diagram Kelola Mata Pelajaran

5. *Sequence Diagram* Kelola Data Pertanyaan

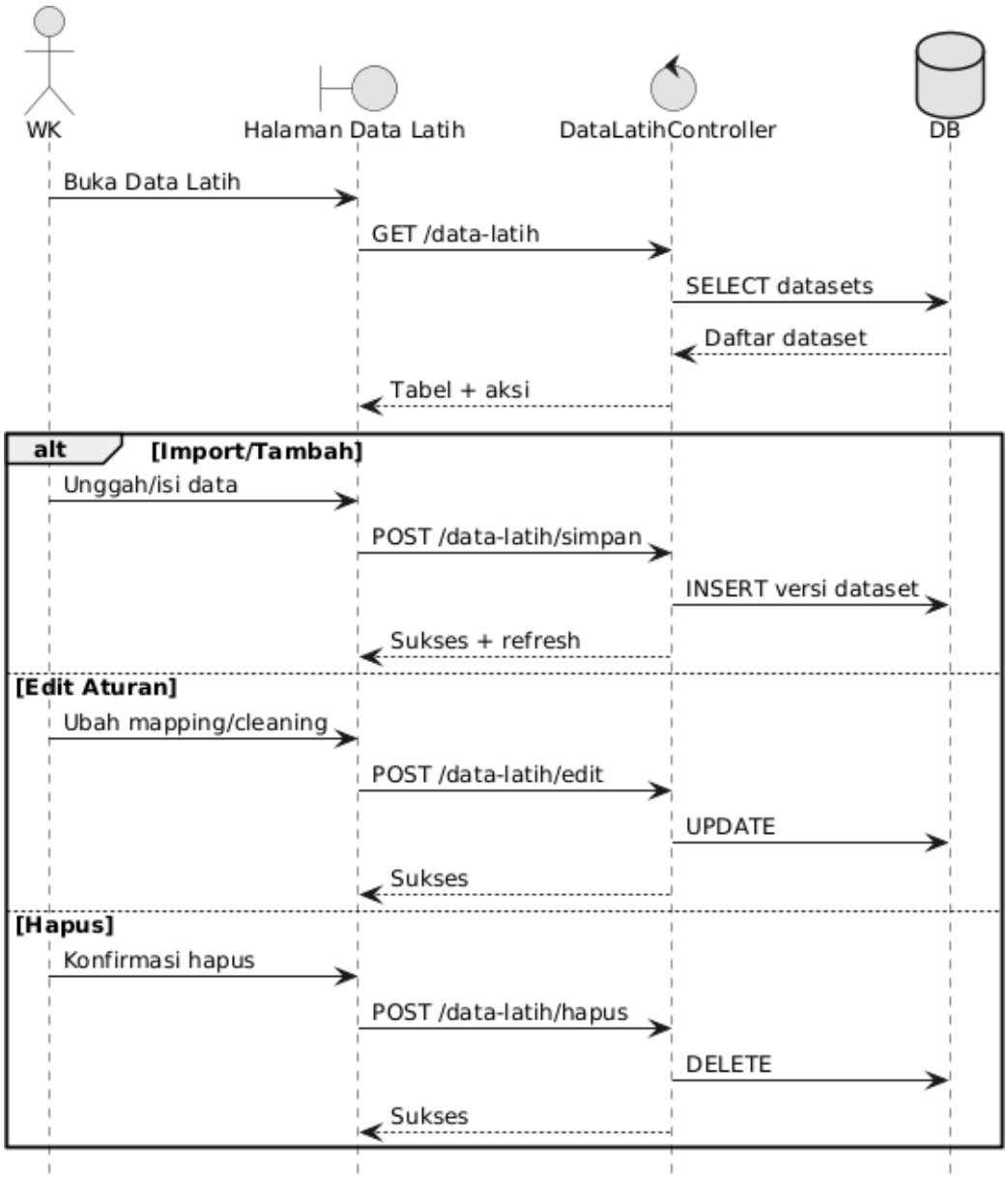
Dalam diagram dibawah ini, Wali Kelas menambah/mengubah /menghapus pertanyaan dan sistem memperbarui DB serta menampilkan daftar terbaru.



Gambar 3. 18 *Sequence Diagram* Kelola Data Pertanyaan

6. *Sequence Diagram* Kelola Data Latih

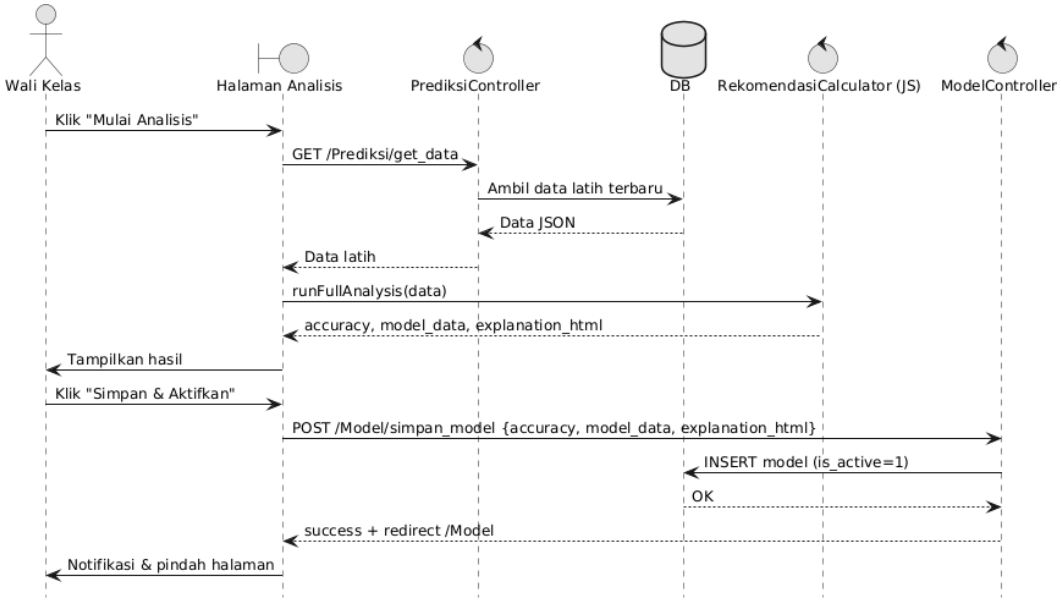
Dalam diagram dibawah ini, Wali Kelas mengelola dataset latih; controller melakukan versioning/validasi dan menyimpan ke DB.



Gambar 3. 19 Sequence Diagram Kelola Data Latih

7. *Sequence Diagram* Analisis dan Latih Model

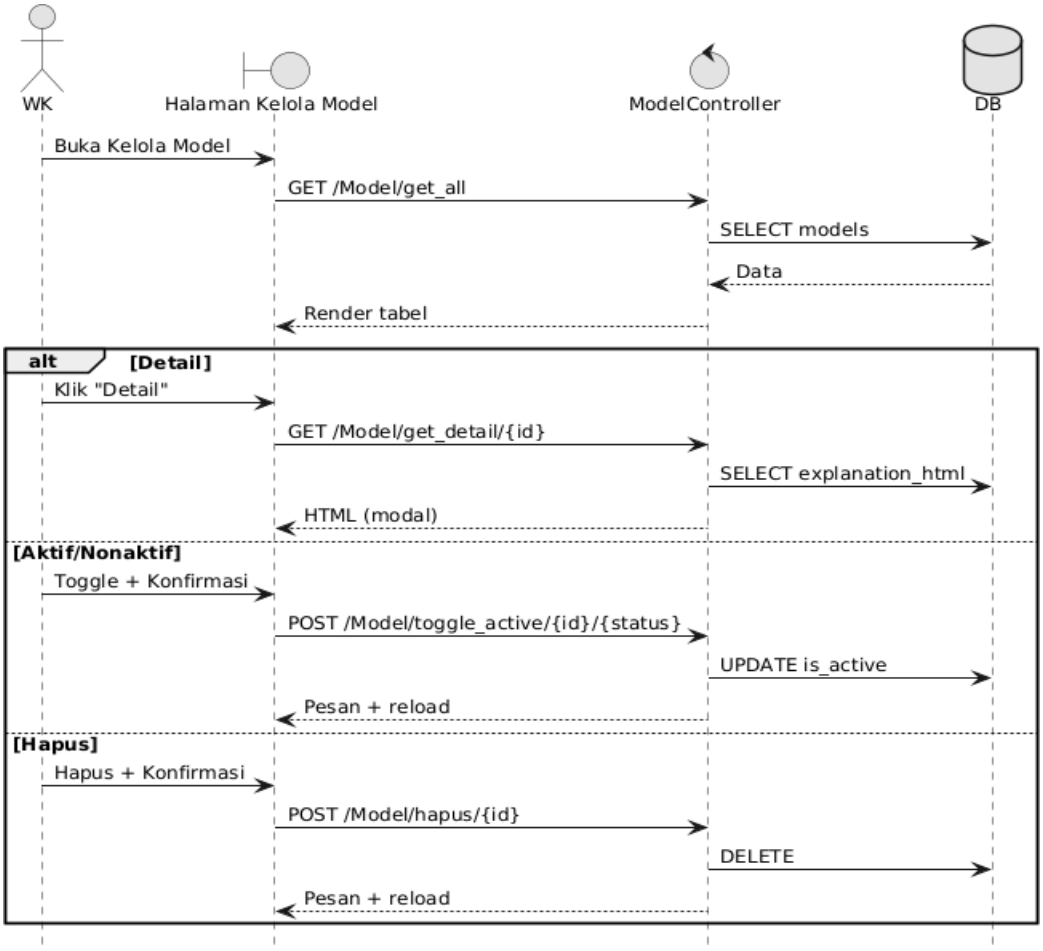
Dalam diagram dibawah ini, UI mengambil data latih, menjalankan analisis di browser (akurasi, model_data, explanation_html), lalu menyimpan & mengaktifkan model.



Gambar 3. 20 Sequence Diagram Analisis dan Latih Model

8. *Sequence Diagram* Kelola Data Model

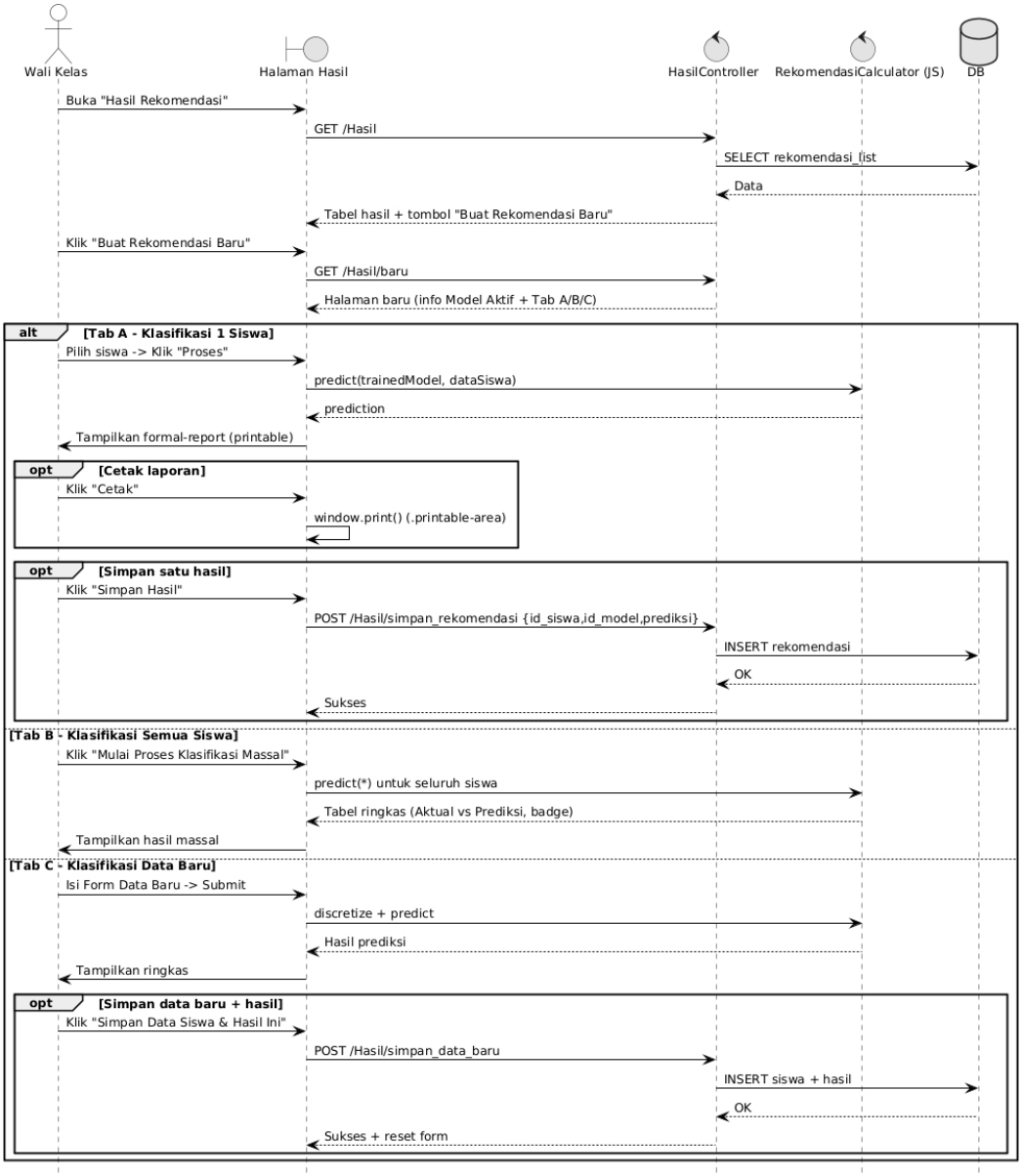
Dalam *Sequence Diagram* dibawah ini, Wali Kelas melihat daftar model, membuka detail, mengaktifkan/nonaktifkan, atau menghapus; DB diperbarui dan UI menampilkan notifikasi.



Gambar 3. 21 Sequence Diagram Manajemen Data Siswa

9. *Sequence Diagram* Hasil Rekomendasi

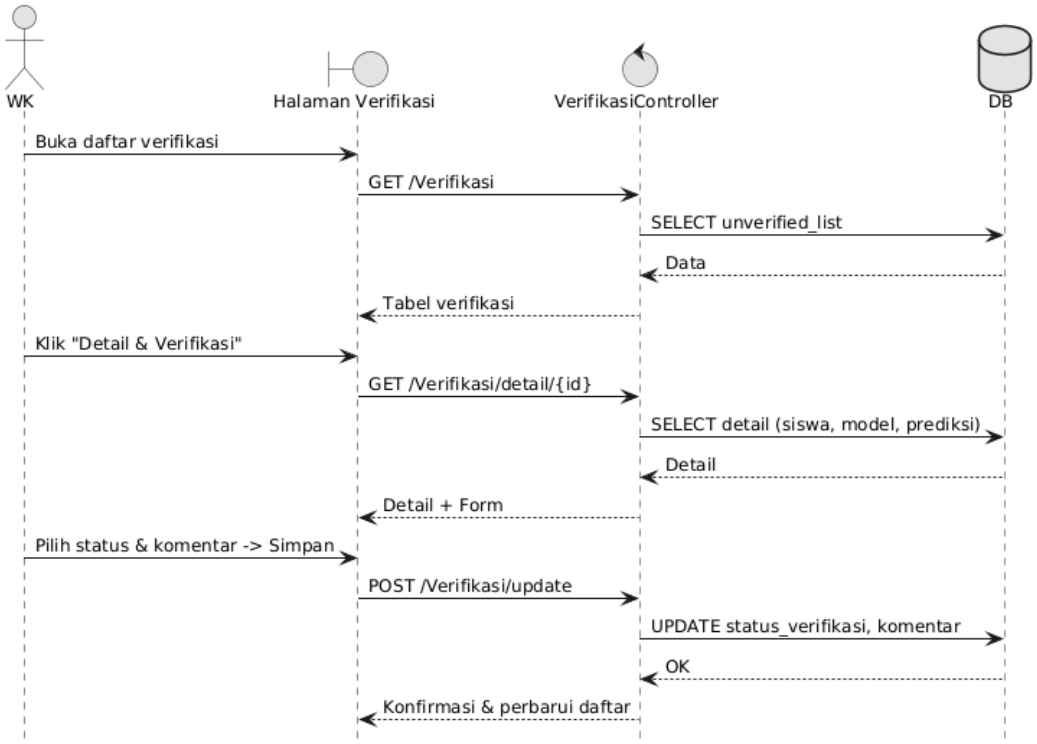
Dalam diagram dibawah ini, Wali Kelas membuka halaman Hasil Rekomendasi untuk melihat daftar yang sudah disimpan, lalu menekan Buat Rekomendasi Baru sehingga sistem menampilkan halaman dengan model aktif dan tiga opsi: Tab A (pilih satu siswa, jalankan prediksi di browser, tampilkan laporan yang bisa dicetak, lalu opsional Simpan Hasil ke database), Tab B (jalankan prediksi massal untuk semua siswa dan tampilkan tabel ringkas kesesuaian), atau Tab C (isi form data siswa baru, sistem memprediksi dan opsional Simpan Data Siswa & Hasil). Setelah selesai, pengguna dapat kembali ke daftar untuk melihat status terbaru.



Gambar 3. 22 *Sequence Diagram* Hasil Rekomendasi

10. *Sequence Diagram* Verifikasi Hasil

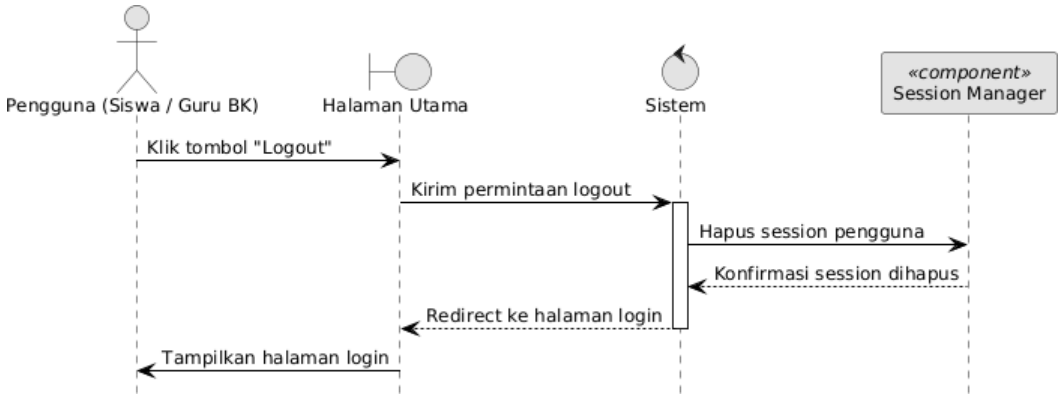
Dalam diagram dibawah ini, Wali Kelas meninjau detail hasil rekomendasi, memilih status verifikasi, menambah komentar, dan sistem menyimpannya lalu memperbarui daftar.



Gambar 3. 23 *Sequence Diagram* Verifikasi Hasil

11. *Sequence Diagram Logout*

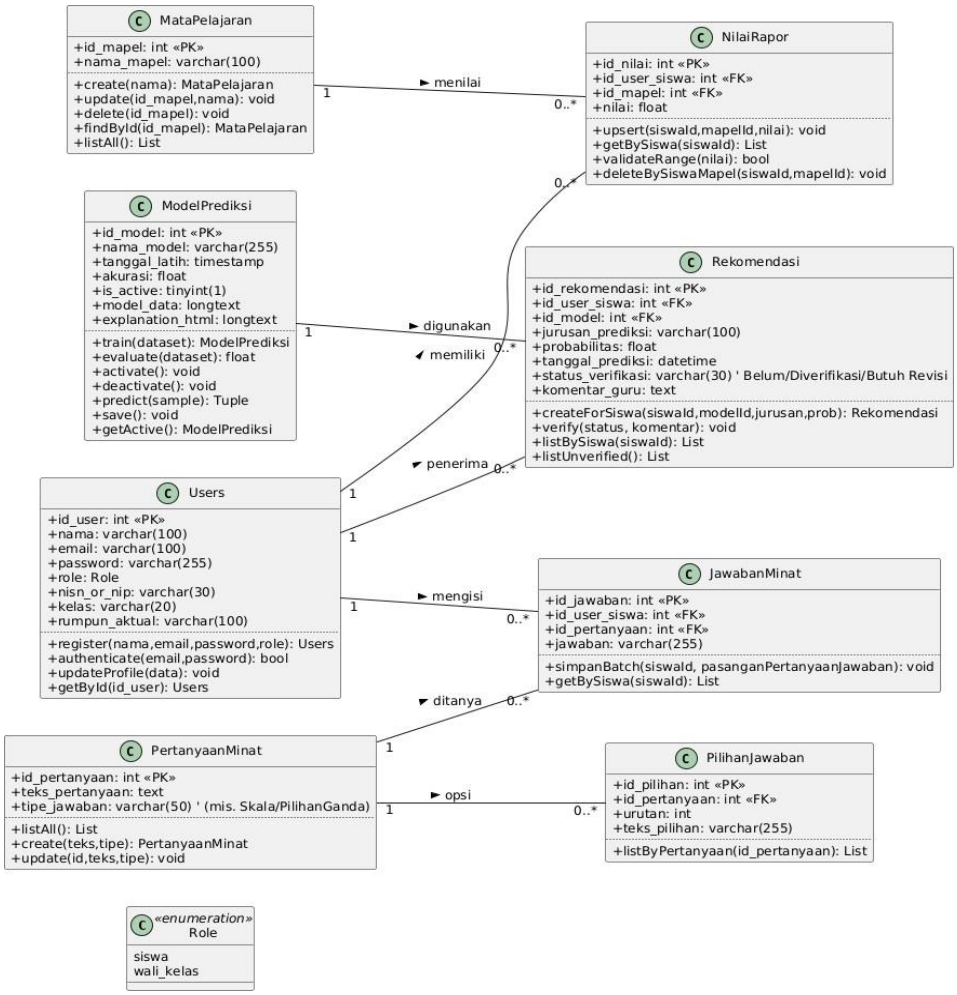
Diagram ini menunjukkan proses logout sederhana untuk Siswa atau Wali Kelas. Pengguna mengklik tombol logout, yang memicu sistem untuk menghapus sesi pengguna melalui "Session Manager". Setelah sesi dihapus, sistem mengarahkan pengguna kembali ke halaman login utama.



Gambar 3. 24 *Sequence Diagram Logout*

3.7.4. *Class Diagram*

Class diagram ini menggambarkan sistem di mana pengguna (siswa/wali_kelas) menjadi pusat. tiap siswa punya Nilai Rapor per Mata Pelajaran dan jawaban kuesioner (Jawaban Minat) atas Pertanyaan Minat (dengan Pilihan Jawaban). Data tersebut melatih dan memberi input ke Model Prediksi yang dapat diaktifkan dan dipakai untuk predict jurusan. Hasilnya disimpan sebagai Rekomendasi per siswa lengkap dengan probabilitas, waktu, status verifikasi, dan komentar dari wali kelas.



Gambar 3. 25 Rancangan *Class Diagram*

3.7.5. Perancangan Basis Data

Perancangan basis data ini bertujuan untuk mendukung semua kebutuhan fungsional sistem dalam menyimpan, mengelola, dan memproses informasi dari pengguna. Struktur ini dirancang untuk menangani data secara terintegrasi, akurat, dan efisien, terutama dalam proses input nilai, tes minat, pengolahan model, serta penyajian hasil rekomendasi jurusan.

1. Rancangan Tabel *users*

   Tabel ini berfungsi untuk menyimpan akun dan profil dasar pengguna. Peran dibatasi pada siswa dan wali_kelas agar kontrol akses sederhana dan jelas.

Tabel 3. 2 Rancangan Tabel *Users*

Field	Type	Key	Keterangan
id_user	INT AUTO_INCREMENT	**PK**	Identitas unik pengguna
nama	VARCHAR(100)	–	Nama lengkap
email	VARCHAR(100)	**UNIQUE**	Email login, tidak boleh ganda
password	VARCHAR(255)	–	Hash kata sandi (BCrypt/Argon2)
role	ENUM('siswa','wali_kelas')	–	Peran pengguna
nisn_or_nip	VARCHAR(30)	UNIQUE NULL	NISN (siswa) / NIP (wali)
kelas	VARCHAR(20)	–	Kelas siswa (opsional)
rumpun_aktual	VARCHAR(100)	–	Jurusan/rumpun saat ini (jika ada)

2. Rancangan Tabel Mata Pelajaran

   Tabel ini digunakan untuk master data nama mata pelajaran yang menjadi referensi nilai rapor.

Tabel 3. 3 Rancangan Tabel Mata Pelajaran

Field	Type	Key	Keterangan
id_mapel	INT AUTO_INCREMENT	**PK**	ID mapel
nama_mapel	VARCHAR(100)	UNIQUE	Nama unik mata pelajaran

3. Rancangan Tabel Nilai Rapor

Tabel ini berfungsi untuk Menyimpan nilai rapor per siswa per mapel sebagai fitur akademik utama untuk pemodelan.

Tabel 3. 4 Rancangan Tabel Nilai Rapor

Field	Type	Key	Keterangan
id_nilai	INT AUTO_INCREMENT	**PK**	ID nilai
id_user_siswa	INT	**FK→users.id_user**	Siswa pemilik nilai
id_mapel	INT	**FK→mata_pelajaran.id_mapel**	Mapel terkait
nilai	DECIMAL(5,2)	–	Rentang 0–100 (beri CHECK)

4. Rancangan Tabel Pertanyaan Minat

Tabel ini digunakan sebagai tempat untuk menyimpan pertanyaan kuesioner minat/gaya belajar yang menjadi fitur non-akademik.

Tabel 3. 5 Rancangan Tabel Pertanyaan Minat

Field	Type	Key	Keterangan
id_pertanyaan	INT AUTO_INCREMENT	**PK**	ID pertanyaan
teks_pertanyaan	TEXT	–	Isi pertanyaan
tipe_jawaban	VARCHAR(50)	–	Mis. Skala, PilihanGanda, gaya_belajar

5. Rancangan Tabel Pilihan Jawaban

Tabel ini menyimpan hasil Opsi jawaban untuk pertanyaan minat bertipe pilihan ganda.

Tabel 3. 6 Rancangan Pilihan Jawaban

Field	Type	Key	Keterangan
id_pilihan	INT AUTO_INCREMENT	**PK**	ID opsi
id_pertanyaan	INT	**FK→pertanyaan_minat.id_pertanyaan**	Pertanyaan induk
urutan	INT	–	Urutan tampilan
teks_pilihan	VARCHAR(255)	–	Teks opsi

6. Rancangan Tabel Pilihan Jawaban

Tabel ini menyimpan hasil Opsi jawaban untuk pertanyaan minat bertipe pilihan ganda.

Tabel 3. 6 Rancangan Pilihan Jawaban

Field	Type	Key	Keterangan
id_pilihan	INT AUTO_INCREMENT	**PK**	ID opsi
id_pertanyaan	INT	**FK→pertanyaan_minat.id_pertanyaan**	Pertanyaan induk
urutan	INT	–	Urutan tampilan
teks_pilihan	VARCHAR(255)	–	Teks opsi

7. Rancangan Tabel Jawaban Minat

Tabel ini menyimpan informasi Jawaban kuesioner yang diisi siswa yang menjadi fitur minat untuk training/prediksi.

Tabel 3. 7 Rancangan Tabel Jawaban Minat

Field	Type	Key	Keterangan
id_jawaban	INT AUTO_INCREMENT	**PK**	ID jawaban
id_user_siswa	INT	**FK→users.id_user**	Siswa pengisi
id_pertanyaan	INT	**FK→pertanyaan_minat.id_pertanyaan**	Pertanyaan

8. Rancangan Tabel Model Prediksi

Tabel ini Menyimpan model hasil pelatihan (di browser/JS), metrik, dan penjelasan agar dapat ditinjau kembali.

Tabel 3. 8 Rancangan Tabel Model Prediksi

Field	Type	Key	Keterangan
id_model	INT AUTO_INCREMENT	**PK**	ID model
nama_model	VARCHAR(255)	–	Penamaan versi model
tanggal_latih	DATETIME	–	Waktu pelatihan
akurasi	DECIMAL(5,4)	–	Akurasi evaluasi
is_active	TINYINT(1)	INDEX	Penanda model aktif
model_data	LONGTEXT	–	Payload model (JSON)
explanation_html	LONGTEXT	–	Laporan analitik/MathJax

9. Rancangan Tabel Rekomendasi

Tabel ini berfungsi untuk Menyimpan hasil prediksi jurusan untuk siswa oleh model tertentu, berikut status verifikasi dan komentar.

Gambar 3. 26 Rancangan Tabel Rekomendasi

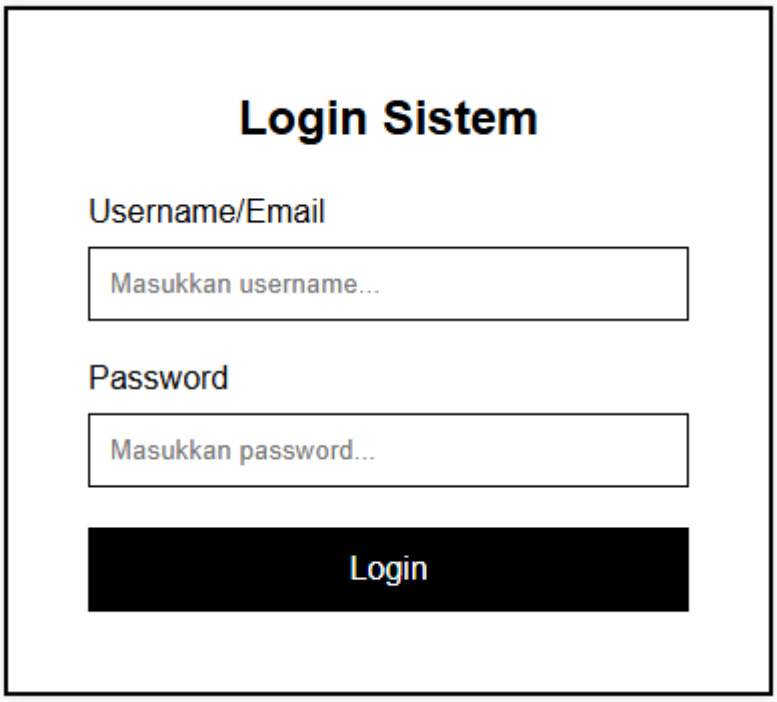
Field	Type	Key	Keterangan
id_rekomendasi	INT AUTO_INCREMENT	**PK**	ID rekomendasi
id_user_siswa	INT	**FK→users.id_user**	Siswa penerima
id_model	INT	**FK→model_prediksi.id_model**	Model yang

			dipakai
jurusan_prediksi	VARCHAR(100)	–	Hasil jurusan
probabilitas	DECIMAL(5,4)	–	Confidence (opsional)
tanggal_prediksi	DATETIME	–	Waktu prediksi
status_verifikasi	ENUM('Belum Diverifikasi','Diverifikasi','Butuh Revisi')	INDEX	Status review
komentar_guru	TEXT	–	Catatan verifikasi

3.8 Rancangan *User* Interface

1. Rancangan Menu Login

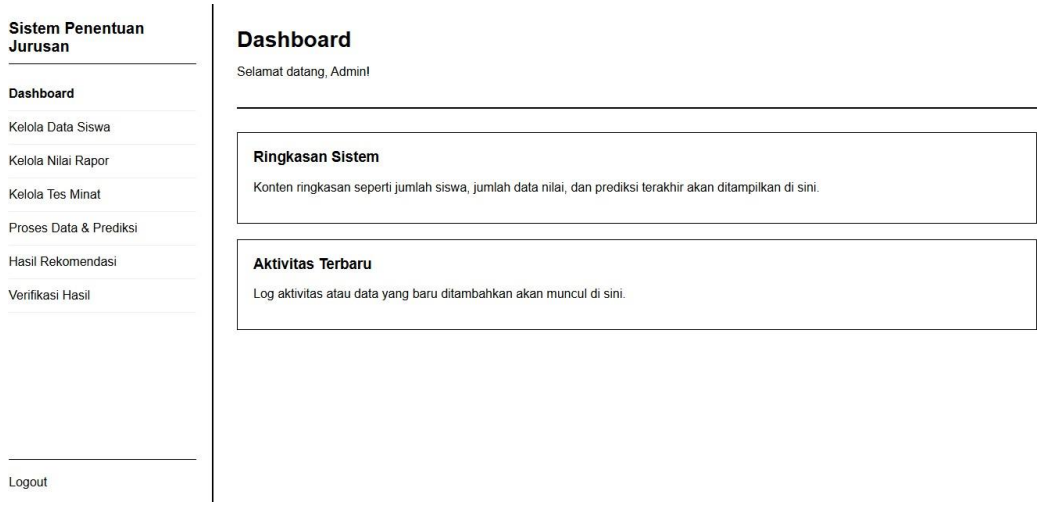
Halaman ini adalah gerbang utama bagi semua pengguna (Wali Kelas, dan Siswa) untuk masuk ke sistem. Rancangan terkait dapat dilihat pada gambar dibawah ini.



Gambar 3. 27 *Interface* Menu Login

2. Rancangan Menu Dashboard (Wali Kelas)

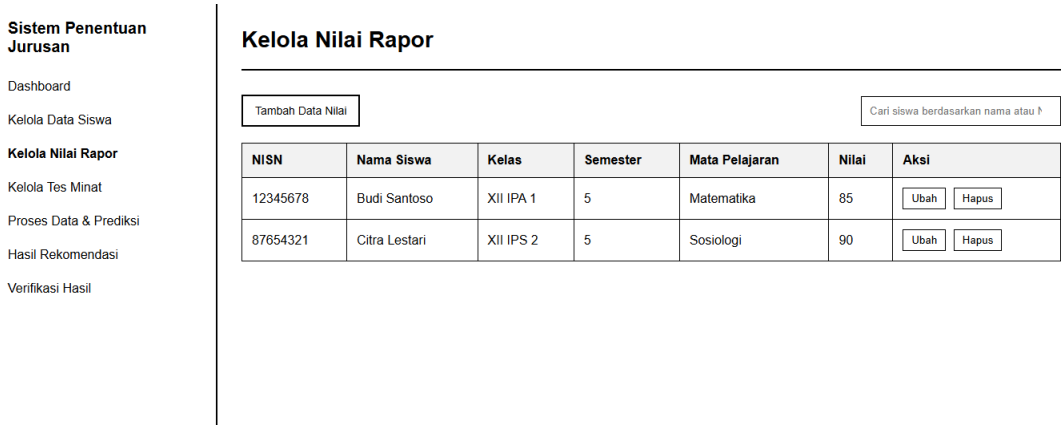
Setelah Wali Kelas akan diarahkan ke dashboard yang berisi menu navigasi di samping dan area konten utama. Dashboard menampilkan ringkasan informasi dan akses ke semua fitur pengelolaan, dapat dilihat pada gambar dibawah ini.



Gambar 3. 28 *Interface Dashboard* Admin dan Wali Kelas

3. Rancangan Menu Kelola Nilai Rapor

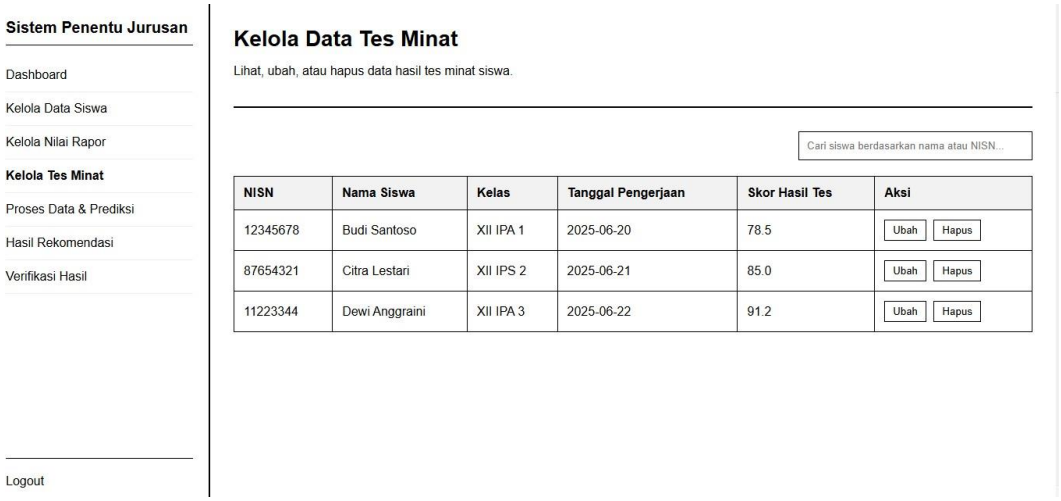
Halaman ini digunakan oleh Wali Kelas untuk menambah, mengubah, dan menghapus data nilai rapor siswa. Dapat dilihat pada gambar dibawah ini.



Gambar 3. 29 *Interface* Kelola Nilai Rapor

4. Rancangan Menu Kelola Tes Minat

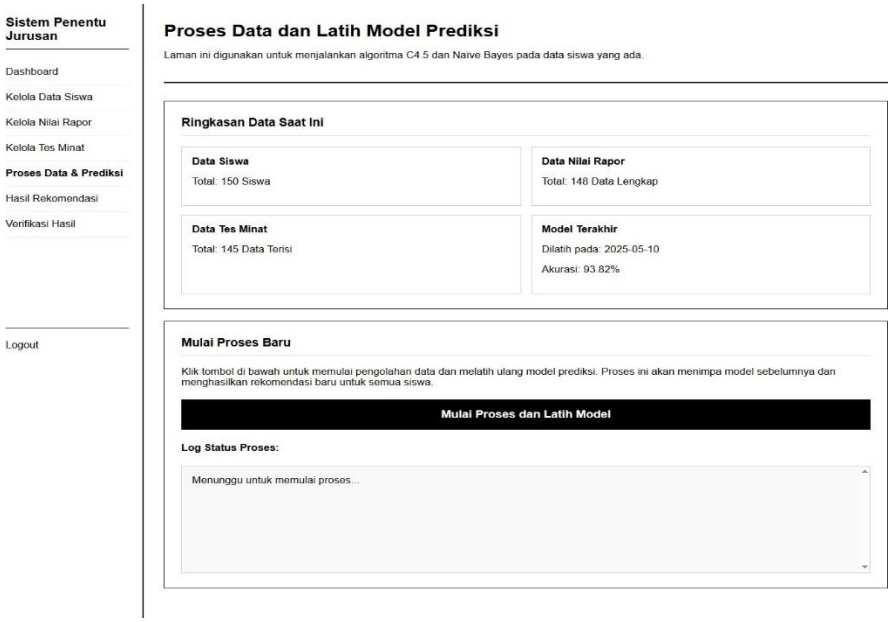
Halaman ini menampilkan daftar hasil tes minat dari semua siswa. Wali Kelas dapat mencari siswa tertentu, melihat hasil skor mereka, serta melakukan perubahan atau penghapusan data jika diperlukan.



Gambar 3. 30 Interface Menu Kelola Data Tes Minat

5. Rancangan Menu Proses Data dan Prediksi

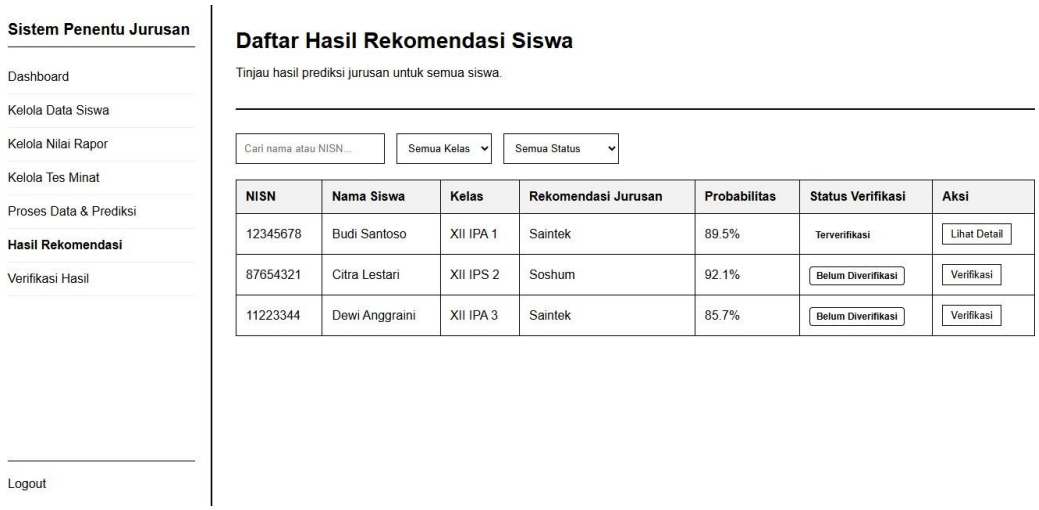
Halaman ini merupakan pusat dari fungsionalitas sistem, yang khusus digunakan oleh Wali Kelas Di sini, Wali Kelas dapat memulai proses pengolahan data mulai dari seleksi fitur menggunakan C4.5 hingga klasifikasi dengan Naïve Bayes untuk menghasilkan rekomendasi jurusan bagi seluruh siswa.



Gambar 3. 31 *Interface* Menu Proses Data dan Latih Model Prediksi

6. Rancangan Menu Hasil Rekomendasi

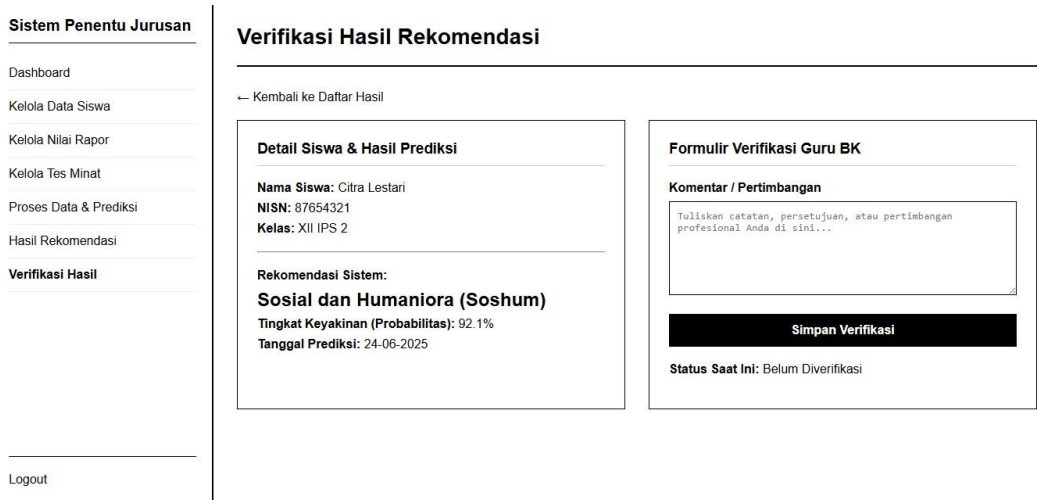
Halaman ini menampilkan tabel berisi ringkasan hasil rekomendasi dari seluruh siswa. Wali Kelas dapat dengan cepat melihat jurusan yang disarankan, tingkat keyakinan prediksi, dan status verifikasi.



Gambar 3. 32 *Interface* Menu Hasil Rekomendasi

7. Rancangan Menu Verifikasi Hasil

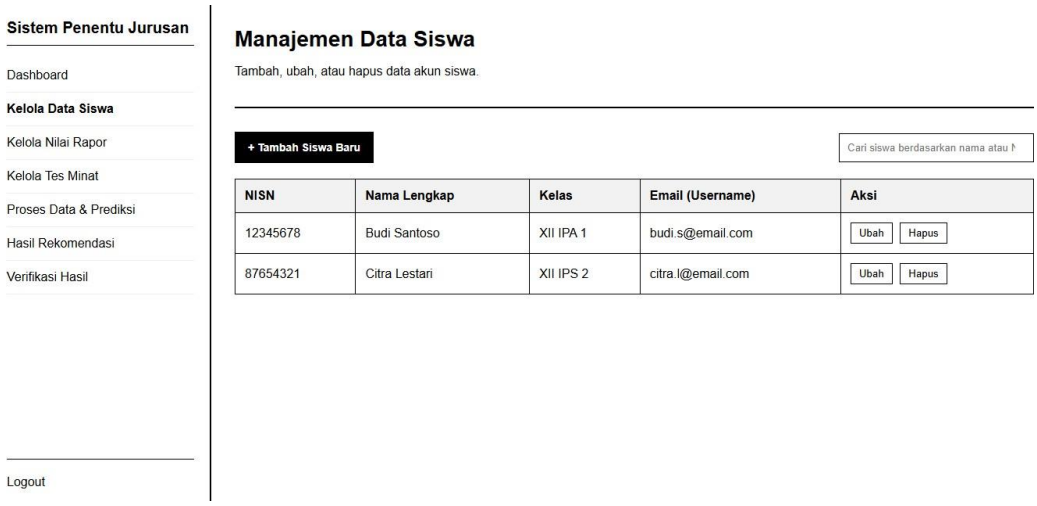
Halaman ini menampilkan detail hasil rekomendasi untuk satu siswa. Wali Kelas dapat meninjau informasi lengkap, kemudian memberikan komentar dan menyimpan status verifikasi.



Gambar 3. 33 *Interface* Menu Verifikasi Hasil

8. Rancangan Menu Manajemen Data Siswa

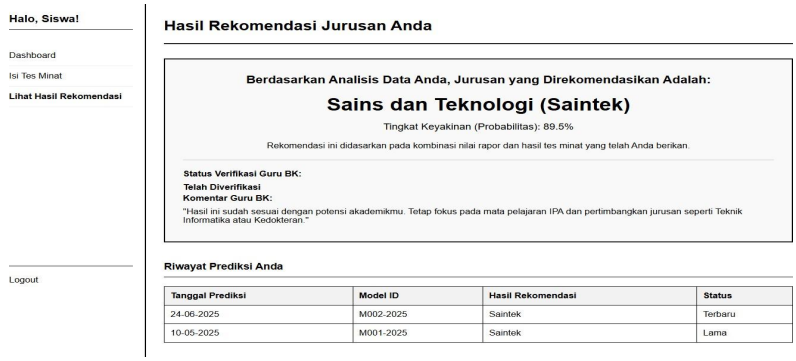
Halaman ini berisi tabel lengkap data siswa yang terdaftar di sistem. Dilengkapi dengan fitur pencarian dan tombol aksi untuk menambah, mengubah, dan menghapus data



Gambar 3. 34 Interface Menu Manajemen Data Siswa

9. Rancangan Lihat Hasil Dan Riwayat Prediksi (Siswa)

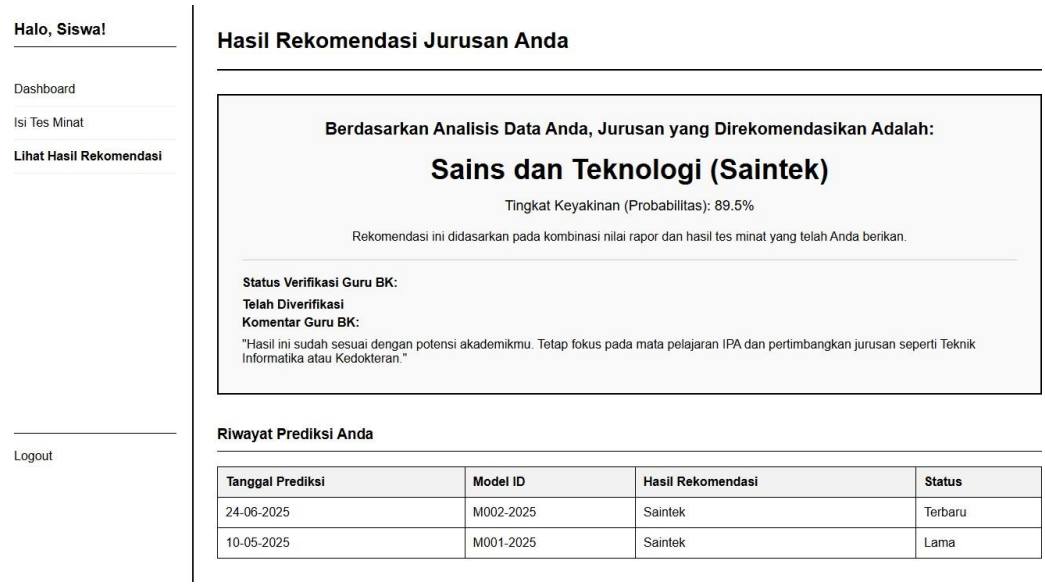
Halaman ini merupakan tujuan akhir bagi siswa setelah semua data diproses oleh sistem. Di sini, siswa dapat melihat secara jelas jurusan apa yang direkomendasikan untuknya, lengkap dengan tingkat keyakinan dari prediksi tersebut. Selain itu, halaman ini juga menampilkan status verifikasi dan komentar dari Wali Kelas serta riwayat prediksi sebelumnya.



Gambar 3. 35 Interface Lihat Hasil

10. Rancangan Lihat Hasil Dan Riwayat Prediksi (Siswa)

Halaman ini merupakan tujuan akhir bagi siswa setelah semua data diproses oleh sistem. Di sini, siswa dapat melihat secara jelas jurusan apa yang direkomendasikan untuknya, lengkap dengan tingkat keyakinan dari prediksi tersebut. Selain itu, halaman ini juga menampilkan status verifikasi dan komentar dari Wali Kelas serta riwayat prediksi sebelumnya.



Gambar 3. 35 Interface Lihat Hasil

BAB IV
HASIL DAN PEMBAHASAN

4.1. Perhitungan Manual

Untuk studi kasus perhitungan manual ini, dilakukan terhadap 160 total data siswa yang tersebar dalam 4 kelas. Data kemudian dibagi menjadi 80% untuk data latih dan 20 persen untuk data uji dengan pembagian sebagai berikut :

a) 128 Data Latih : Untuk membangun dan melatih model C4.5

b) 32 Data Uji : Untuk menguji dan mengevaluasi

Performa model Berikut merupakan distribusi keseluruhan untuk keseluruhan data

Gambar 4. 1 Distribusi Keseluruhan Data

Gambar 4. 2 Distribusi Kelas Pada Data Latih

Gambar 4. 3 Distribusi Kelas Pada Data Uji

Seluruh atribut numerik (nilai rapor/indikator minat) didiskretisasi tiga tingkat (*Tinggi*, *Sedang*, *Rendah*). Diskretisasi memudahkan pembentukan tabel

frekuensi dan probabilitas bersyarat pada Naïve Bayes serta mempermudah evaluasi *entropy* pada perhitungan *Information Gain*. Berikut merupakan atruannya untuk setiap fitur numerik.

Tabel 4. 1 Tabel Diskretisasi Nilai

Rentang Nilai	Kategori
≥ 90	Tinggi
80 ≤ nilai < 90	Sedang
< 80	Rendah

Setelah proses diatas, berikut merupakan beberapa dari data latih dan data uji

Tabel 4. 2 Contoh Tabel Data Latih

Nama Siswa	NISN	Matematika	Ekonomi	Minat Logika	Biologi	Fisika	Minat Interaksi Sosial & Masyarakat	Pancasila	Minat Bisnis & Kewirausahaan	Kimia	Sejarah	Rumpun Rekomendasi
Aditya Ruspramono	0076425998	Rendah	Rendah	Suka	Rendah	Sedang	Suka	Tinggi	Biasa Saja	Rendah	Rendah	Sosial, Hukum & Pem

												erintahan
AIRIN LAURA MARCEL A MANURUNG	00773 61567	Sedang	Rendah	Sangat Suka	Sedang	Tinggi	Sangat Tidak Suka	Rendah	Suka	Rendah	Rendah	RTI
ALYA WAHYUN I	00796 84156	Tinggi	Rendah	Suka	Sedang	Sedang	Biasa Saja	Rendah	Sangat Tidak Suka	Rendah	Rendah	RTI
ANNISA BILKHIST I	00747 57022	Sedang	Rendah	Biasa Saja	Sedang	Sedang	Sangat Tidak Suka	Rendah	Biasa Saja	Sedang	Tinggi	Pendidikan, Bahasa & Budaya
AULIA NURMUZ DALIFAH	00729 22167	Rendah	Rendah	Biasa Saja	Sedang	Rendah	Sangat Suka	Sedang	Sangat Suka	Rendah	Rendah	Sosial, Hukum & Pemerintaha

												n
Azhar Rafii	0078929440	Rendah	Rendah	Biasa Saja	Sedang	Rendah	Biasa Saja	Sedang	Tidak Suka	Sedang	Tinggi	Pendidikan, Bahasa &

Selanjutnya, dilakukan seleksi fitur dengan Logika C4.5, Seleksi fitur merupakan tahap penting dalam proses pembangunan model klasifikasi karena tidak semua atribut pada dataset memiliki kontribusi yang sama terhadap hasil prediksi. Pada penelitian ini digunakan pendekatan logika C4.5, khususnya perhitungan *Information Gain*, untuk menentukan fitur mana yang paling relevan. *Information Gain* mengukur seberapa besar penurunan entropi (ketidakpastian) yang dihasilkan ketika data dibagi berdasarkan suatu atribut. Dengan kata lain, semakin besar nilai *gain* sebuah fitur, semakin besar pula pengaruh fitur tersebut dalam membedakan kelas target. Proses ini dilakukan dengan menghitung entropi total dari dataset, kemudian membandingkannya dengan entropi dari subset data yang terbentuk berdasarkan nilai-nilai fitur yang bersangkutan.

Hasil dari perhitungan *Information Gain* selanjutnya digunakan untuk mengurutkan seluruh fitur dari yang paling tinggi hingga terendah. Fitur dengan *gain* tertinggi dipandang sebagai fitur yang paling informatif karena mampu memberikan kejelasan terbesar dalam memprediksi kelas target. Dalam penelitian ini, hanya sepuluh fitur teratas yang dipilih untuk digunakan pada tahap pelatihan Naïve Bayes, sehingga model yang dihasilkan lebih sederhana namun tetap mempertahankan informasi penting. Strategi ini juga membantu mengurangi kompleksitas komputasi dan menghindari *overfitting*, karena hanya atribut yang benar-benar relevan yang dipertahankan dalam proses klasifikasi.

Langkah pertama adalah perhitungan Entropi total pada data latih

menggunakan rumus dibawah ini :

$$Entropy\ (S) = -\sum_{c} p(c)\ log_2 p(c)$$

$Entropy\ (Keseluruhan)$

$$-1\ x \log \frac{34}{128} + -\frac{22}{128} x \log \frac{22}{128} + -\frac{8}{128} x \log \frac{26}{128} + -\frac{29}{128} x \log \frac{29}{128} + -\frac{17}{128} x \log \frac{17}{128}$$

$$Entropy\ (Keseluruhan)$$

$$= (-0.2656\ x - 1.19125) + (-0.1719\ x - 2.5406)$$

$$+ (-0.2031\ x - 2.2996) + \ (-0.2266\ x - 2.1420)$$

$$+ (-0.1327\ x - 2.9125) = 2.2839$$

Berikut merupakan perhitungan *Entropy* dan *Information Gain* untuk setiap fitur

1. Fitur 1 : Nilai Matematika

   Langkah 1 – Hitung Entropi untuk setiap nilai dari fitur Matematika

   a) Untuk Nilai Sedang (61 data dari total 128 data)

Tabel 4. 3 Tabel perhitungan Entropy Matematika Nilai Sedang

Kelas	Jumlah	p	$-p \cdot \log_2(p)$
Ekonomi & Bisnis	19	0.3115	0.5242
Pendidikan, Bahasa & Budaya	9	0.1475	0.4073
Sosial, Hukum & Pemerintahan	14	0.2295	0.4873
RTI	8	0.1311	0.3844
Kesehatan & Ilmu Alam	11	0.1803	0.4456
Total Entropi			**2.2488**

   b) Untuk Nilai Tinggi (36 data dari total 128 data)

Tabel 4. 4 Tabel perhitungan Entropy Matematika Nilai Tinggi

Kelas	Jumlah	p	$-p \cdot \log_2(p)$
Ekonomi & Bisnis	15	0.4167	0.5263
RTI	21	0.5833	0.4536
Total Entropi			**0.9799**

c) Untuk Nilai Rendah (31 data dari total 128 data)

Tabel 4. 5 Tabel perhitungan Entropy Matematika Nilai Rendah

Kelas	**Jumlah**	**p**	**$-p \cdot \log_2(p)$**
Sosial, Hukum & Pemerintahan	12	0.3871	0.5300
Pendidikan, Bahasa & Budaya	13	0.4194	0.5258
Kesehatan & Ilmu Alam	6	0.1935	0.4586
Total Entropi (SRendah)			**1.5144**

Langkah 2 – Hitung *Information Gain*

$$Gain\,(Matematika) = Entropy\,(S) - \sum_{v \in a} \frac{|S_v|}{|S|} Entropy\,(S_v)$$

$Gain\,(Matematika)$

$= 2.2839 - [(\frac{6}{128} x\ 1.2488) + (\frac{36}{128} x\ 0.9799) + (\frac{28}{128} x\ 1.5144)]$

$Gain\,(Matematika) = 2.2839 - [1.0717 + 0.2756 + 0.3668]$

$Gain\,(Matematika) = 0.5698$

Langkah perhitungan yang sama dilakukan pada ke 17 fitur sehingga menghasilkan urutan *Information Gain* seperti pada tabel dibawah ini

Tabel 4. 6 Tabel urutan *Information Gain*

Peringkat	Fitur	Information Gain
1	Matematika	0.5698
2	Ekonomi	0.5158
3	Minat pada Logika & Pemecahan Masalah	0.4983
4	Biologi	0.4747
5	Fisika	0.4742
6	Minat pada Interaksi Sosial & Masyarakat	0.4716
7	Pancasila	0.4311
8	Minat pada Bisnis & Kewirausahaan	0.4246
9	Kimia	0.4223
10	Sejarah	0.3725
11	Bahasa Indonesia	0.3548
12	Geografi	0.3532
13	Minat pada Seni & Kreativitas	0.3387
14	Minat pada Sains & Penelitian	0.3360
15	Bahasa Inggris	0.3342
16	Sosiologi	0.3294
17	Gaya Belajar yang Paling Disukai	0.0689

Untuk membangun model yang efisien dan fokus pada atribut paling relevan, maka dipilih 10 fitur teratas untuk digunakan dalam tahap pelatihan Naive

Bayes. Pemilihan ini bertujuan untuk mengurangi noise dari fitur yang kurang informatif dan mencegah *overfitting*.

Berdasarkan tahap perhitungan seleksi pada Berdasarkan tahap seleksi fitur sebelumnya, proses pelatihan Naive Bayes akan menggunakan 10 fitur yitu, Matematika, Ekonomi, Minat pada Logika & Pemecahan Masalah, Biologi, Fisik, Minat pada Interaksi Sosial & Masyarakat, Pancasila, Minat pada Bisnis & Kewirausahaan, Kimia, Sejarah

$P(Ekonomi \& Bisnis)$ = $\frac{34}{128}$ = 0.2656

$P(Kesehatan\ dan\ Ilmu\ Alam)$ = $\frac{17}{128}$ = 0.1328

$P(Pendidikan, Bahasa, dan\ Budaya)$ = $\frac{22}{128}$ = 0.1719

$P(RTI)$ = $\frac{29}{128}$ = 0.2266

$P(Sosial, Hukum, Pemerintahan)$ = 26 = 0.2031

Setelah itu dihitung probabilitas kemunculan suatu fitur jika diketahui kelasnya (*likelihood)*, berikut perhitungan di Fitur Matematika

Perhitungan Matematika untuk Fitur Sedang

Tabel 4. 7 Perhitungan Likelihood Fitur Matematika Nilai Sedang

Kelas (H)	**count(Sedang, H)**	**count(H)**	**k**	**Rumus (Laplace)**	**P(Sedang|H)**
Ekonomi & Bisnis	19	34	3	(19+1) / (34+3) = 20/37	**0.5405**
Kesehatan & Ilmu Alam	11	17	3	(11+1) / (17+3) = 12/20	**0.6000**
Pendidikan, Bahasa & Budaya	9	22	3	(9+1) / (22+3) = 10/25	**0.4000**
RTI	8	29	3	(8+1) / (29+3) = 9/32	**0.2813**
Sosial, Hukum & Pemerintahan	14	26	3	(14+1) / (26+3) = 15/29	**0.5172**

Perhitungan Matematika untuk Fitur Tinggi

Tabel 4. 8 Perhitungan Likelihood Fitur Matematika Nilai Tinggi

Kelas (H)	count(Tinggi, H)	count(H)	k	Rumus (Laplace)	P(Tinggi|H)
Ekonomi & Bisnis	15	34	3	(15+1) / (34+3) = 16/37	**0.4324**
Kesehatan & Ilmu Alam	0	17	3	(0+1) / (17+3) = 1/20	**0.0500**
Pendidikan, Bahasa & Budaya	0	22	3	(0+1) / (22+3) = 1/25	**0.0400**
RTI	21	29	3	(21+1) / (29+3) = 22/32	**0.6875**
Sosial, Hukum & Pemerintahan	0	26	3	(0+1) / (26+3) = 1/29	**0.0345**

Tabel 4. 9 Perhitungan Likelihood Fitur Matematika Nilai Rendah

Kelas (H)	count(Rendah, H)	count(H)	k	Rumus (Laplace)	P(Rendah|H)
Ekonomi & Bisnis	0	34	3	(0+1) / (34+3) = 1/37	**0.0270**
Kesehatan & Ilmu Alam	6	17	3	(6+1) / (17+3) = 7/20	**0.3500**
Pendidikan, Bahasa & Budaya	13	22	3	(13+1) / (22+3) = 14/25	**0.5600**
RTI	0	29	3	(0+1) / (29+3) = 1/32	**0.0313**
Sosial, Hukum & Pemerintahan	12	26	3	(12+1) / (26+3) = 13/29	**0.4483**

Perhitungan dilakukan di semua 10 fitur dilakukan perhitungan probabilitas posterior untuk menentukan rumpun jurusan siswa, berikut merupakan contoh perhitungan untuk siswa atas nama Diajeng Intan Suriamanda dengan data sebagai berikut :

Nama Siswa: DIAJENG INTAN SURIAMANDA NISN: 0073475701

Kelas: XII MIA-2

1. Matematika: 92
2. Bahasa Indonesia: 94
3. Bahasa Inggris: 82
4. Fisika: 97
5. Kimia: 82

6. Biologi: 98
7. Ekonomi: 8
8. Sosiologi: 75
9. Geografi: 81
10. Sejarah: 78
11. Pendidikan Kewarganegaraan: 75
12. Minat terhadap Logika & Teknologi: Tidak Suka
13. Minat terhadap Sains & Alam: Sangat Tidak Suka
14. Minat terhadap Sosial & Masyarakat: Biasa Saja
15. Minat terhadap Bisnis & Ekonomi: Suka
16. Minat terhadap Bahasa & Seni: Sangat Suka
17. Gaya Belajar Dominan: Visual (Gambar & Video)
18. Rumpun Rekomendasi: Pendidikan, Bahasa & Budaya

Data numerik diubah menjadi diskrit menjadi seperti dibawah ini : Nama Siswa: DIAJENG INTAN SURIAMANDA
NISN:
0073475701
Kelas: XII MIA-2

1. Matematika: 92 → Tinggi
2. Bahasa Indonesia: 94 → Tinggi
3. Bahasa Inggris: 82 → Sedang
4. Fisika: 97 → Tinggi
5. Kimia: 82 → Sedang
6. Biologi: 98 → Tinggi
7. Ekonomi: 84 → Sedang
8. Sosiologi: 75 → Rendah

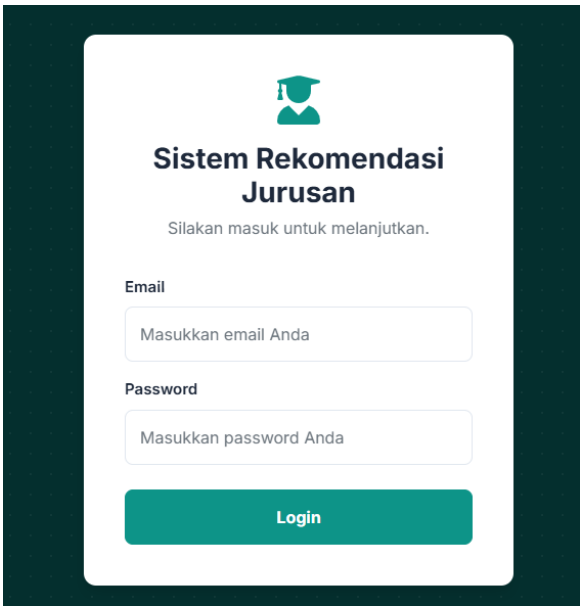
9. Geografi: 81 → Sedang
10. Sejarah: 78 → Rendah
11. Pendidikan Kewarganegaraan: 75 → Rendah
12. Minat terhadap Logika & Teknologi: Tidak Suka
13. Minat terhadap Sains & Alam: Sangat Tidak Suka
14. Minat terhadap Sosial & Masyarakat: Biasa Saja
15. Minat terhadap Bisnis & Ekonomi: Suka
16. Minat terhadap Bahasa & Seni: Sangat Suka
17. Gaya Belajar Dominan: Visual (Gambar & Video)
18. Rumpun Rekomendasi: Pendidikan, Bahasa & Budaya

Perhitungan Posterior dihitung dengan menjumlahkan log prior dan log likelihood semua fitur. Hasil perhitungan skor log menunjukkan:
Ekonomi & Bisnis = -13.3821 Kesehatan & Ilmu Alam = 14.2806
Pendidikan, Bahasa & Budaya = -
12.7760 RTI = -13.0945
Sosial, Hukum & Pemerintahan = -12.3239
Kelas dengan skor tertinggi (log paling besar atau paling mendekati nol) adalah Sosial, Hukum & Pemerintahan. Dengan demikian, model merekomendasikan siswa tersebut masuk rumpun Sosial, Hukum & Pemerintahan. Perhitungan dilakukan ke seluruh data, sehingga menghasilkan label yang dapat dilihat dalam implementasi sistem.

4.2. Implementasi Basis Web

1. Tampilan Halaman *Login*

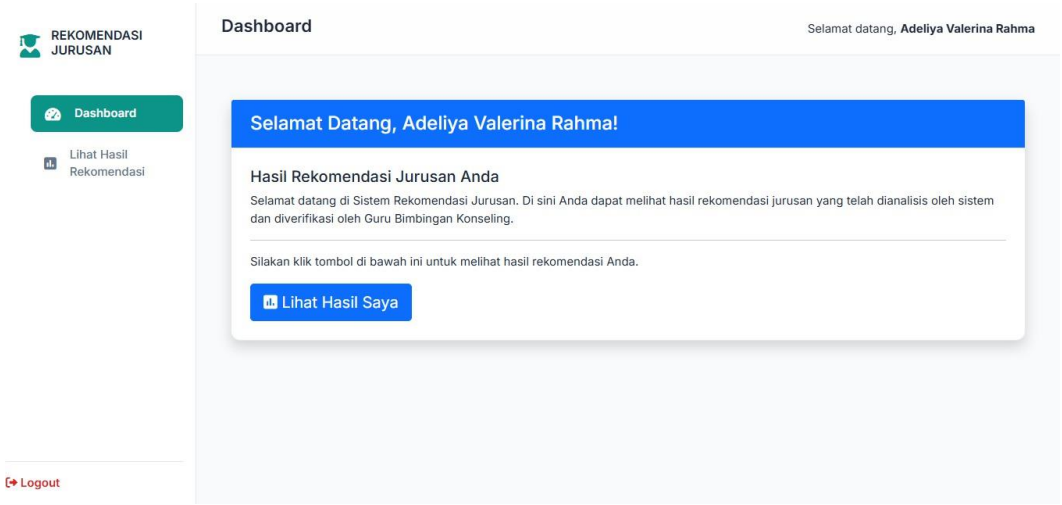
Halaman ini berfungsi sebagai tempat masuk ke sistem, di mana pengguna harus memasukkan email dan *password* untuk mendapatkan akses



Gambar 4. 5 Tampilan Halaman Dashboard (Wali Kelas)

2. Tampilan Halaman *Dashboard* (Siswa)

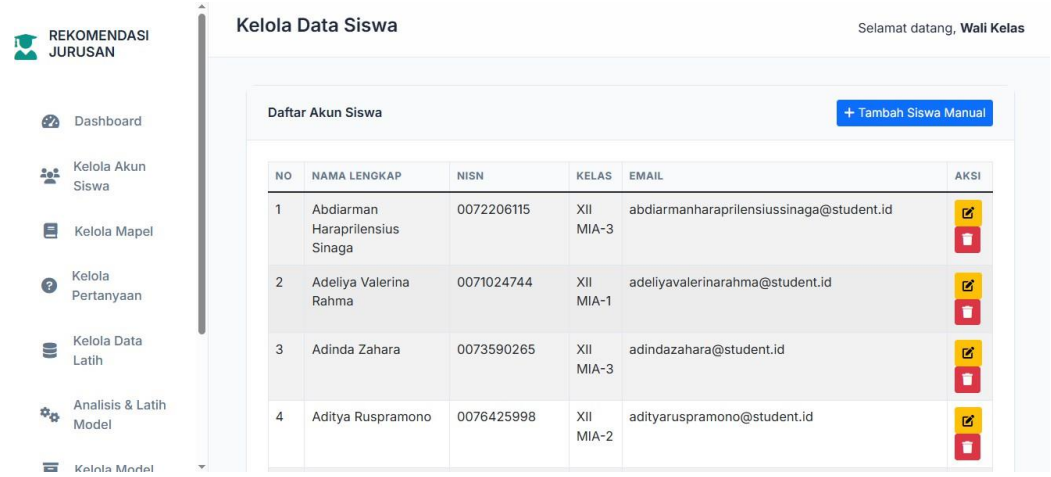
Setelah *login*, siswa akan melihat halaman sambutan sederhana yang mengarahkan mereka untuk melihat hasil rekomendasi jurusan yang telah diproses.



Gambar 4. 6 Tampilan Halaman Dashboard (Siswa)

3. Tampilan Halaman Kelola Akun Siswa

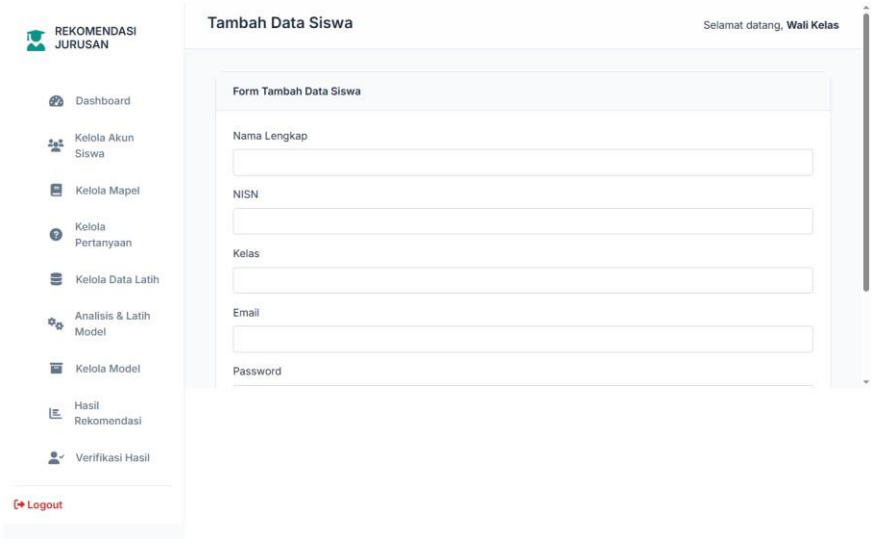
Halaman ini menampilkan daftar seluruh akun siswa yang terdaftar dalam tabel. Wali kelas bisa menambahkan siswa baru secara manual atau mengubah dan menghapus data siswa yang sudah ada.



Gambar 4. 7 Tampilan Halaman Kelola Akun Siswa

4. Tampilan Halaman Tambah & Edit Siswa

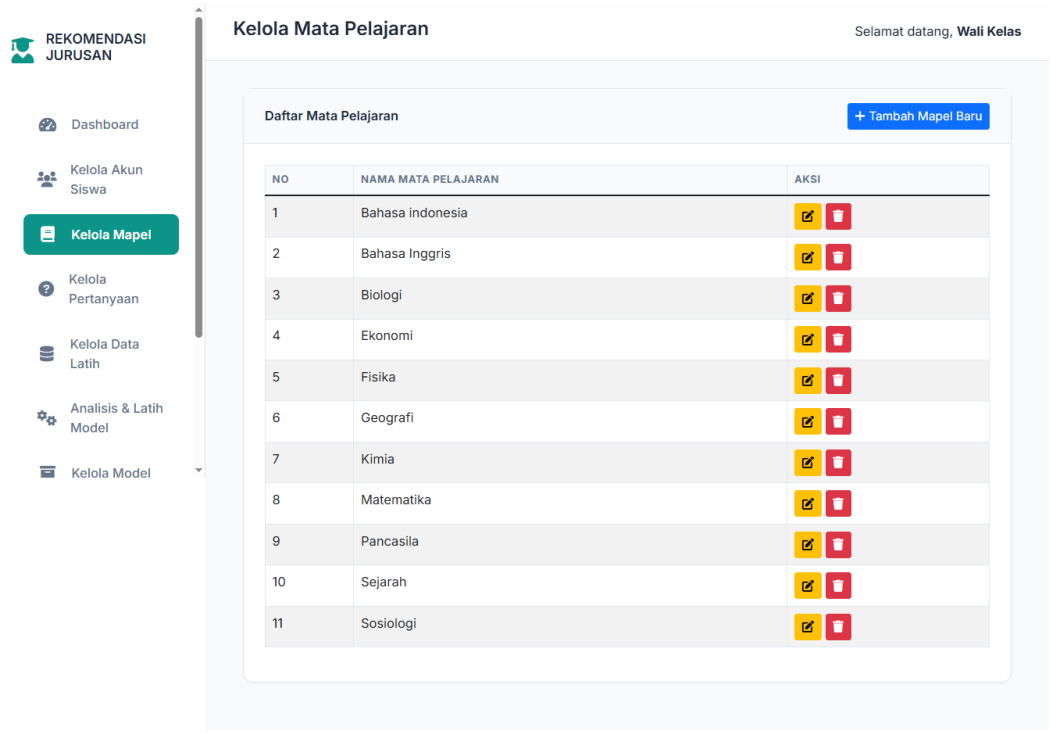
Halaman ini berfungsi menambahkan siswa baru secara manual atau mengubah dan data siswa yang sudah ada.



Gambar 4. 8 Tampilan Halaman Tambah & Edit Siswa

5. Tampilan Halaman Kelola Mata Pelajaran

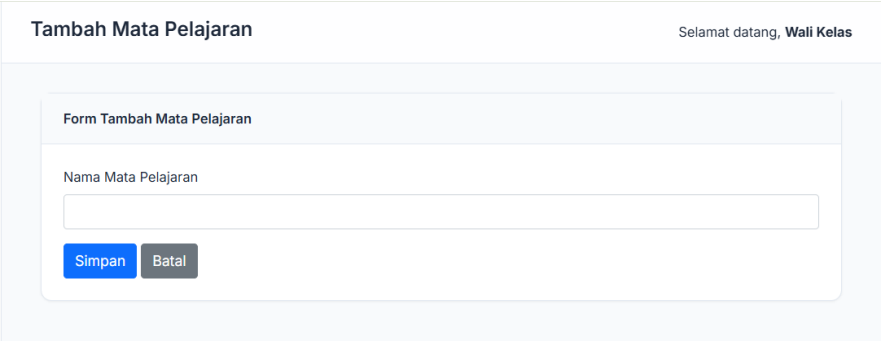
Fitur ini menyajikan daftar mata pelajaran yang digunakan dalam penilaian dan menyediakan formulir sederhana untuk menambah mata pelajaran baru ke dalam sistem.



Gambar 4. 9 Tampilan Halaman Data Mata Pelajaran

6. Tampilan Halaman Tambah Edit Mata Pelajaran

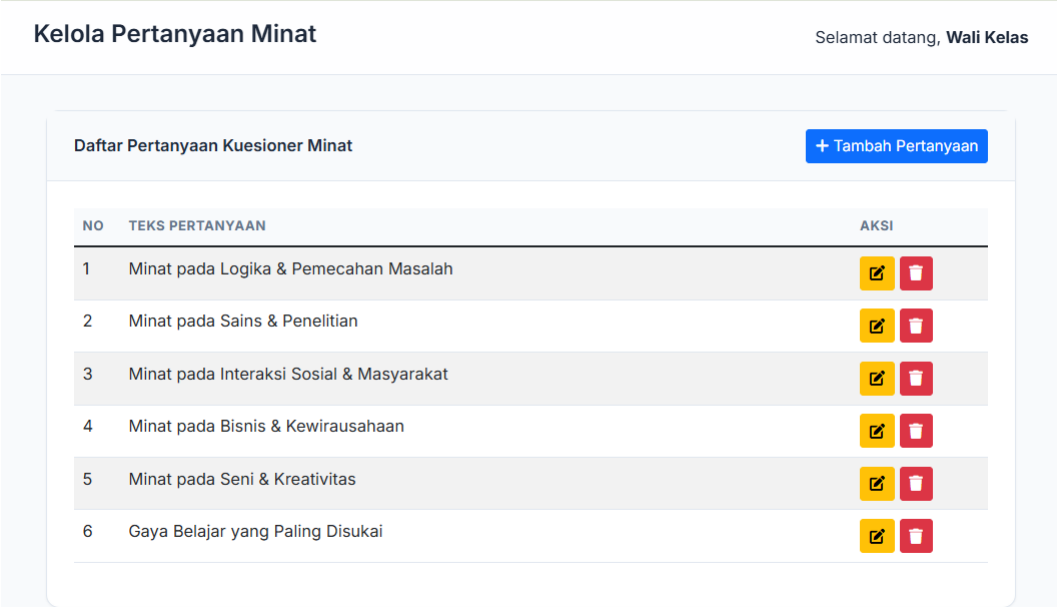
Halaman ini berfungsi menambahkan mata pelajaran baru secara manual atau mengubah dan data mata pelajaran yang sudah ada.



Gambar 4. 10 Tampilan Tambah & Edit Mata Pelajar

7. Tampilan Halaman Kelola Pertanyaan Minat

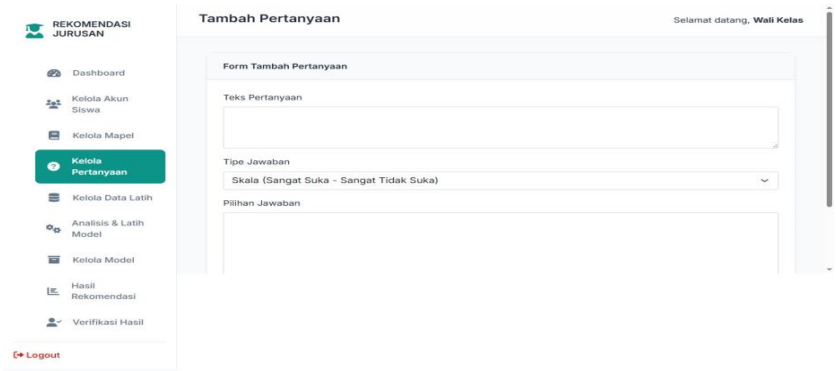
Pada Halaman ini, wali kelas mengelola bank soal untuk kuesioner minat, termasuk menambahkan pertanyaan baru beserta tipe dan pilihan jawabannya.



Gambar 4. 11 Tampilan Halaman Kelola Pertanyaan

8. Tampilan Halaman Tambah & Edit Pertanyaan

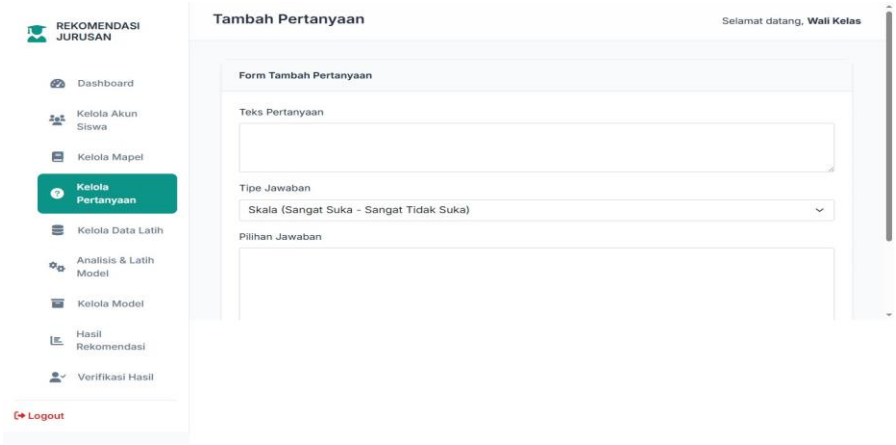
Halaman ini berfungsi menambahkan pertanyaan baru secara manual atau mengubah dan data pertanyaan yang sudah ada.



Gambar 4. 12 Tampilan Halaman Tambah & Edit Pertanyaan

9. Tampilan Halaman Tambah & Edit Pertanyaan

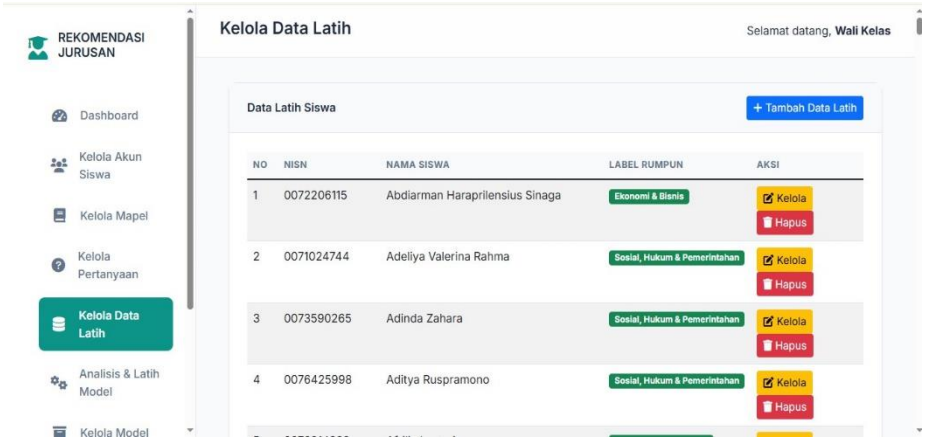
Halaman ini berfungsi menambahkan pertanyaan baru secara manual atau mengubah dan data pertanyaan yang sudah ada.



Gambar 4. 12 Tampilan Halaman Tambah & Edit Pertanyaan

10. Tampilan Halaman Kelola Data Latih

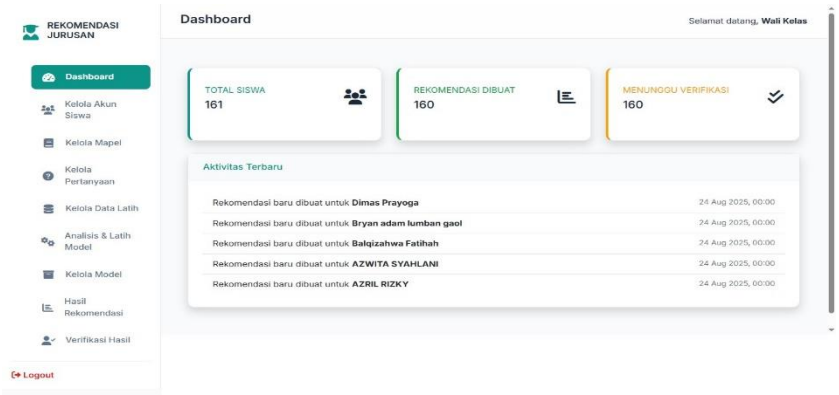
Pada Halaman ini, wali kelas mengelola data latih termasuk menambahkan nilai dan mengisi kuesioner siswa untuk digunakan dalam Analisa menggunakan C45 dan *Naïve Bayes*.



Gambar 4. 4 Tampilan Halaman Login

11. Tampilan Halaman *Dashboard* (Wali Kelas)

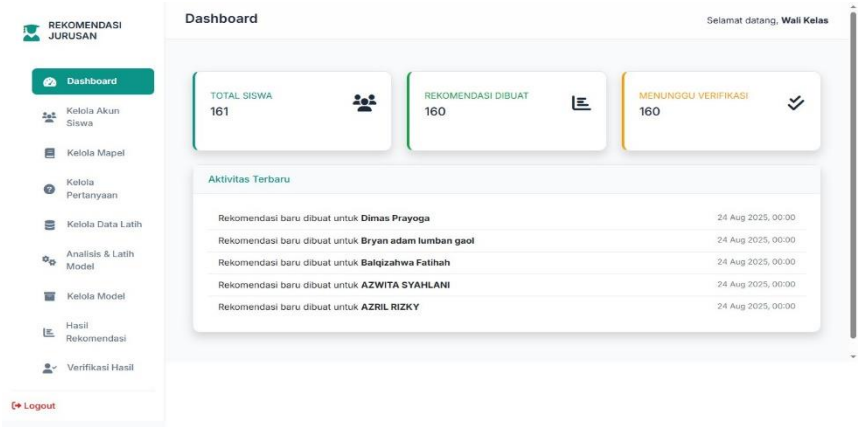
Dasbor ini menyediakan ringkasan data penting bagi wali kelas, seperti total siswa, jumlah rekomendasi yang dibuat, dan daftar aktivitas terbaru.



Gambar 4. 13 Tampilan Halaman Kelola Data Latih

12. Tampilan Halaman Tambah / Kelola Data Latih

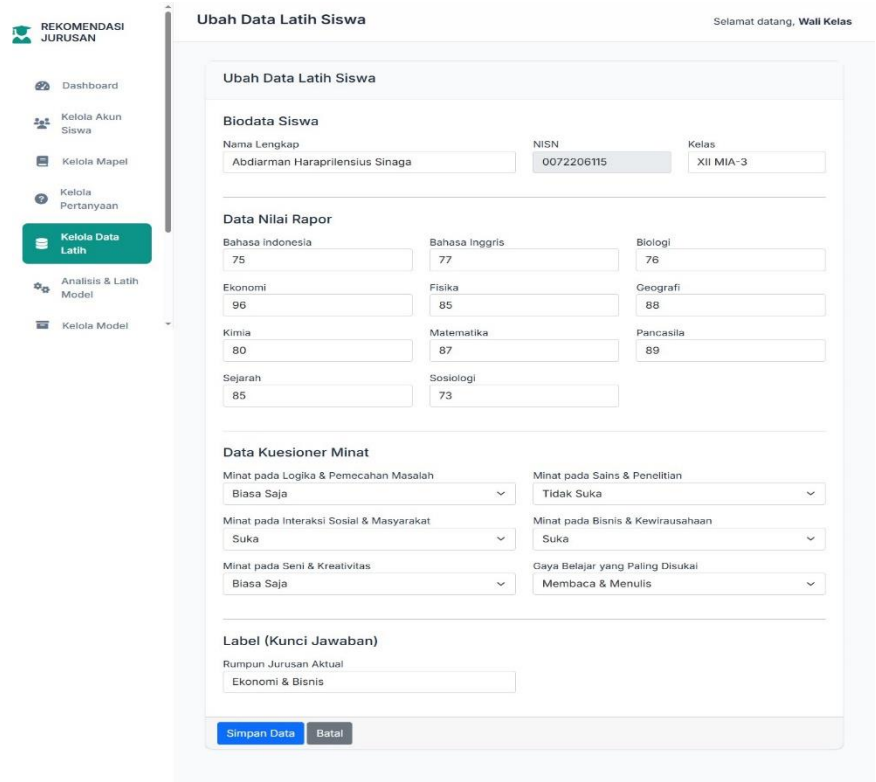
Halaman ini berfungsi sebagai pusat untuk mengelola profil lengkap seorang siswa secara terperinci. Di sini, wali kelas dapat melihat dan mengubah berbagai informasi penting yang akan digunakan sebagai data latih untuk model prediksi, termasuk biodata, nilai rapor per mata pelajaran, jawaban kuesioner minat, hingga menetapkan label jurusan yang sebenarnya (rumpun aktual) sebagai kunci jawaban.



Gambar 4. 13 Tampilan Halaman Kelola Data Latih

13. Tampilan *Halaman* Tambah / Kelola Data Latih

Halaman ini berfungsi sebagai pusat untuk mengelola profil lengkap seorang siswa secara terperinci. Di sini, wali kelas dapat melihat dan mengubah berbagai informasi penting yang akan digunakan sebagai data latih untuk model prediksi, termasuk biodata, nilai rapor per mata pelajaran, jawaban kuesioner minat, hingga menetapkan label jurusan yang sebenarnya (rumpun aktual) sebagai kunci jawaban.



Gambar 4. 14 Tampilan Halaman Kelola Data Latih

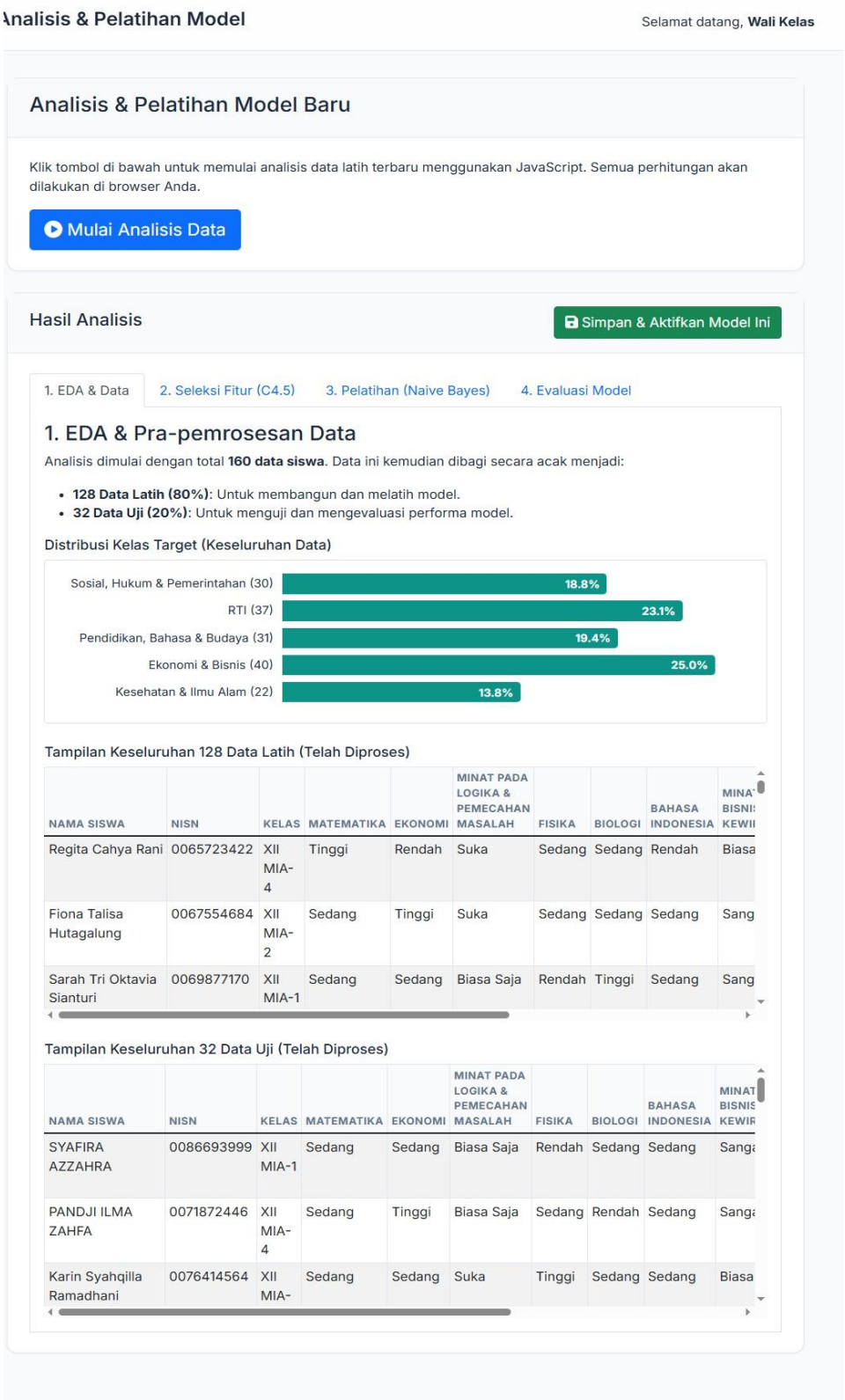
14. Tampilan dan Fungsi Halaman Analisis dan Latih Model

Berikut merupakan alur kerja untuk fitur Analisis & Pelatihan Model berdasarkan urutan gambar yang Anda berikan.

Wali kelas mengawali seluruh proses dengan menekan tombol "Mulai Analisis Data". Tindakan ini memicu sistem untuk mengambil semua data latih siswa yang tersimpan dan memulai serangkaian perhitungan kompleks langsung di *browser* pengguna.

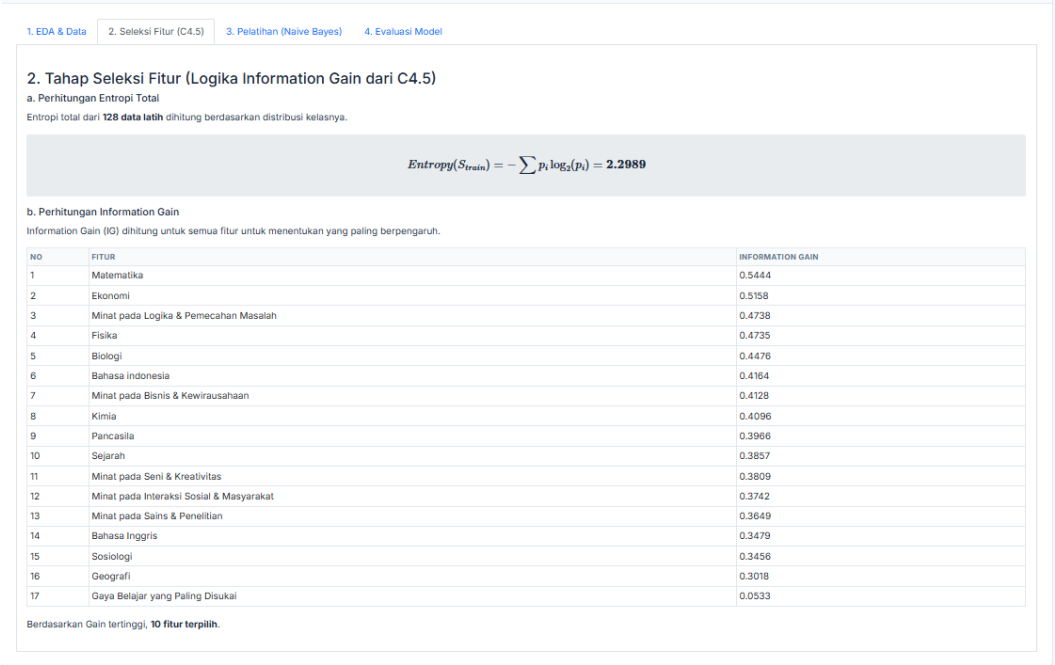
Gambar 4. 15 Memulai Proses Analisis

Setelah proses dimulai, sistem menampilkan tab pertama dari hasil analisis. Di sini, keseluruhan data siswa (dalam contoh ini 160 data) secara otomatis dipecah menjadi 80% data latih (128 data) untuk membangun model dan 20% data uji (32 data) untuk mengevaluasi performanya. Tampilan ini juga menyajikan visualisasi distribusi jurusan di dalam *dataset* dan sampel data yang telah diproses.



Gambar 4. 16 Pra Pemrosesan dan Analisis Data Awal

Selanjutnya, sistem menggunakan logika *Information Gain* dari algoritma C4.5 untuk menentukan fitur (mata pelajaran atau minat) yang paling berpengaruh. Sistem menghitung entropi total dari *dataset* dan kemudian menampilkan skor Information Gain untuk setiap fitur, di mana fitur dengan skor tertinggi adalah yang paling relevan. Berdasarkan skor ini, sistem secara otomatis memilih fitur-fitur terbaik (dalam contoh ini 10 fitur teratas) untuk digunakan pada tahap berikutnya.



Gambar 4. 17 Seleksi Fitur dengan C4.5

Pada tahap selanjutnya, model Naive Bayes dilatih menggunakan fitur-fitur yang sudah terpilih. Sistem secara transparan menunjukkan perhitungannya:

a. Probabilitas Prior: Menghitung peluang awal untuk setiap jurusan berdasarkan frekuensinya di data latih.
b. Probabilitas *Likelihood*: Menunjukkan probabilitas sebuah nilai (misal: nilai 'Tinggi' pada 'Matematika') muncul pada jurusan tertentu.
c. Probabilitas Posterior: Menggabungkan nilai *prior* dan *likelihood* untuk memprediksi jurusan pada sebuah data uji, seperti yang dicontohkan pada siswa Syafira Azzahra.

1. EDA & Data 2. Seleksi Fitur (C4.5) 3. Pelatihan (Naive Bayes) 4. Evaluasi Model

3. Tahap Pelatihan Naïve Bayes

Model Naive Bayes dilatih dengan asumsi independensi antar fitur dan menggunakan **Laplace Smoothing** untuk mengatasi probabilitas nol.

a. Probabilitas Prior

Peluang awal setiap kelas dihitung sebagai berikut:

$$P(H_i) = \frac{\text{Jumlah kasus kelas } H_i}{\text{Total kasus}}$$

KELAS (H)	PERHITUNGAN	PRIOR P(H)
Ekonomi & Bisnis	33 / 128	0.2578
Kesehatan & Ilmu Alam	21 / 128	0.1641
Pendidikan, Bahasa & Budaya	23 / 128	0.1797
RTI	29 / 128	0.2266
Sosial, Hukum & Pemerintahan	22 / 128	0.1719

Gambar 4. 18 Pelatihan *Naive Bayes* - Perhitungan Prior

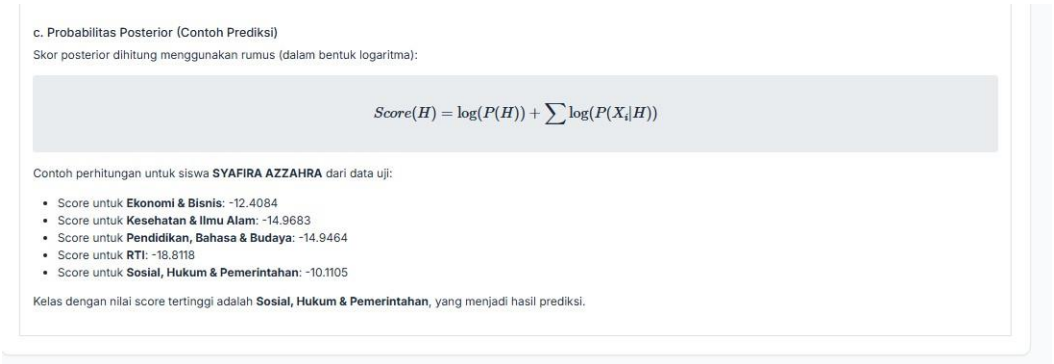
b. Probabilitas Likelihood (Contoh)

Contoh tabel likelihood untuk fitur paling berpengaruh, **'Matematika'**:

$$P(X_j|H_i) = \frac{\text{count}(X_j, H_i) + 1}{\text{count}(H_i) + k}$$

NILAI 'MATEMATIKA'	P(NILAI | EKONOMI & BISNIS)	P(NILAI | KESEHATAN & ILMU ALAM)	P(NILAI | PENDIDIKAN, BAHASA & BUDAYA)	P(NILAI | RTI)	P(NILAI | SOSIAL, HUKUM & PEMERINTAHAN)
Tinggi	0.4722	0.0417	0.0385	0.6563	0.0400
Sedang	0.5000	0.7083	0.4615	0.3125	0.5200
Rendah	0.0278	0.2500	0.5000	0.0313	0.4400

Gambar 4. 19 Pelatihan *Naive Bayes* - Perhitungan *Likelihood*



c. Probabilitas Posterior (Contoh Prediksi)

Skor posterior dihitung menggunakan rumus (dalam bentuk logaritma):

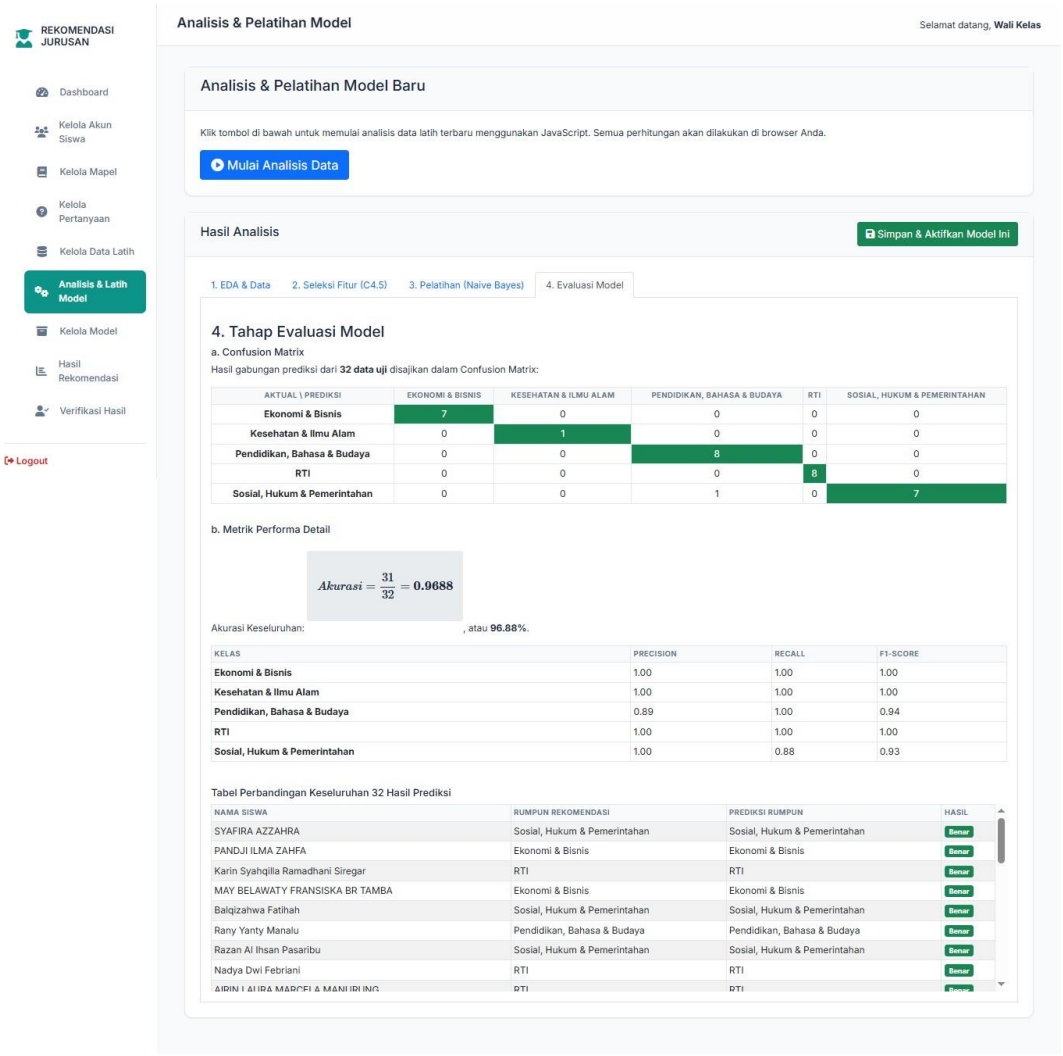
$$Score(H) = \log(P(H)) + \sum \log(P(X_i|H))$$

Contoh perhitungan untuk siswa **SYAFIRA AZZAHRA** dari data uji:

- Score untuk **Ekonomi & Bisnis**: -12.4084
- Score untuk **Kesehatan & Ilmu Alam**: -14.9683
- Score untuk **Pendidikan, Bahasa & Budaya**: -14.9464
- Score untuk **RTI**: -18.8118
- Score untuk **Sosial, Hukum & Pemerintahan**: -10.1105

Kelas dengan nilai score tertinggi adalah **Sosial, Hukum & Pemerintahan**, yang menjadi hasil prediksi.

Gambar 4. 20 Pelatihan *Naive Bayes* - Perhitungan Posterior



Gambar 4. 21 Evaluasi Performa Model

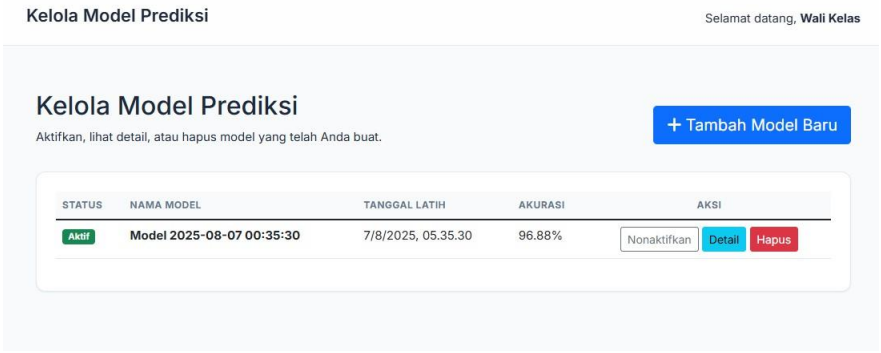
Setelah model selesai dilatih, sistem langsung mengujinya menggunakan data uji yang tadi disisihkan. Hasilnya ditampilkan dalam *Confusion Matrix* untuk melihat keakuratan prediksi per kelas dan metrik performa detail seperti *Precision*, *Recall*, dan *F1-Score*. Akurasi keseluruhan model (dalam contoh ini 96.88%) juga disajikan dengan jelas, diikuti tabel perbandingan antara hasil prediksi dan jurusan yang sebenarnya untuk setiap data uji.

Jika wali kelas puas dengan hasil analisis dan performa model, mereka dapat menyimpannya dengan menekan tombol "Simpan & Aktifkan Model Ini".

Tindakan ini akan menyimpan model cerdas yang baru saja dibuat ke dalam sistem, membuatnya siap digunakan untuk memberikan rekomendasi kepada siswa lain.

1. Tampilan Halaman Kelola Model

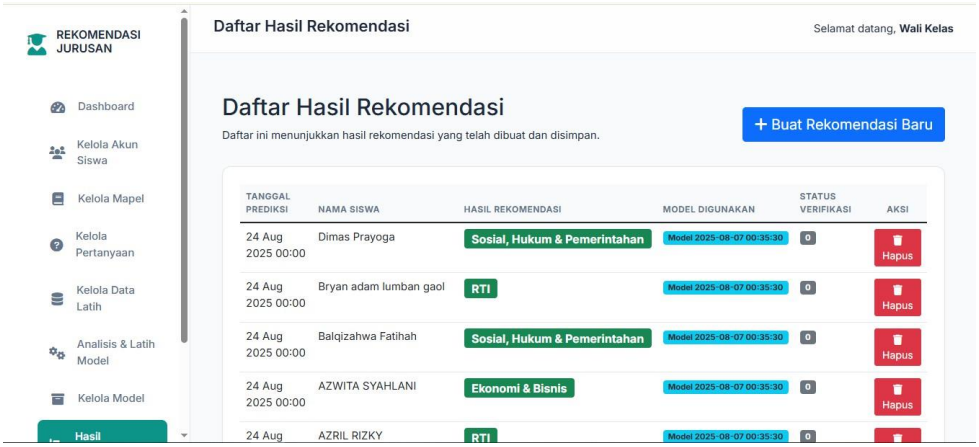
Halaman ini berfungsi sebagai pusat kontrol untuk semua model AI yang telah dibuat. Wali kelas dapat melihat model mana yang sedang aktif, membandingkan akurasi, menonaktifkan, menghapus, atau melihat detail



Gambar 4. 22 Tampilan Halaman Kelola Model

2. Tampilan Halaman Hasil Rekomendasi (Wali Kelas)

Halaman ini menampilkan riwayat semua rekomendasi yang pernah dibuat untuk seluruh siswa. Wali kelas dapat dengan cepat melihat hasil prediksi, model yang digunakan, status verifikasi, dan memiliki opsi untuk menghapus data rekomendasi atau membuat rekomendasi baru.



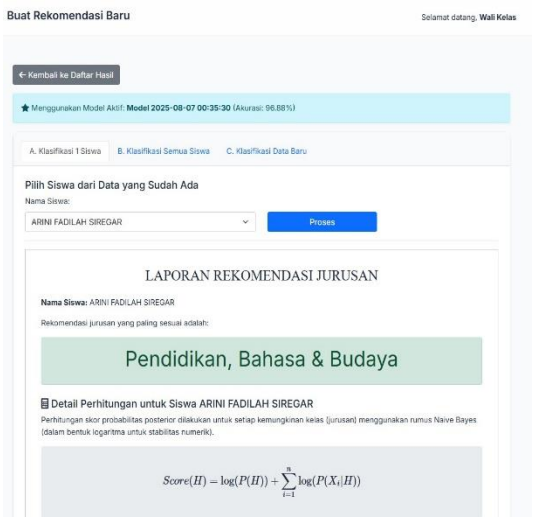
Gambar 4. 23 Tampilan Hasil Rekomendasi

3. Tampilan Halaman Rekomendasi Baru

Halaman ini adalah workspace utama untuk menghasilkan prediksi menggunakan model AI yang aktif. Halaman ini terbagi menjadi tiga tab fungsional:

a) Klasifikasi 1 Siswa

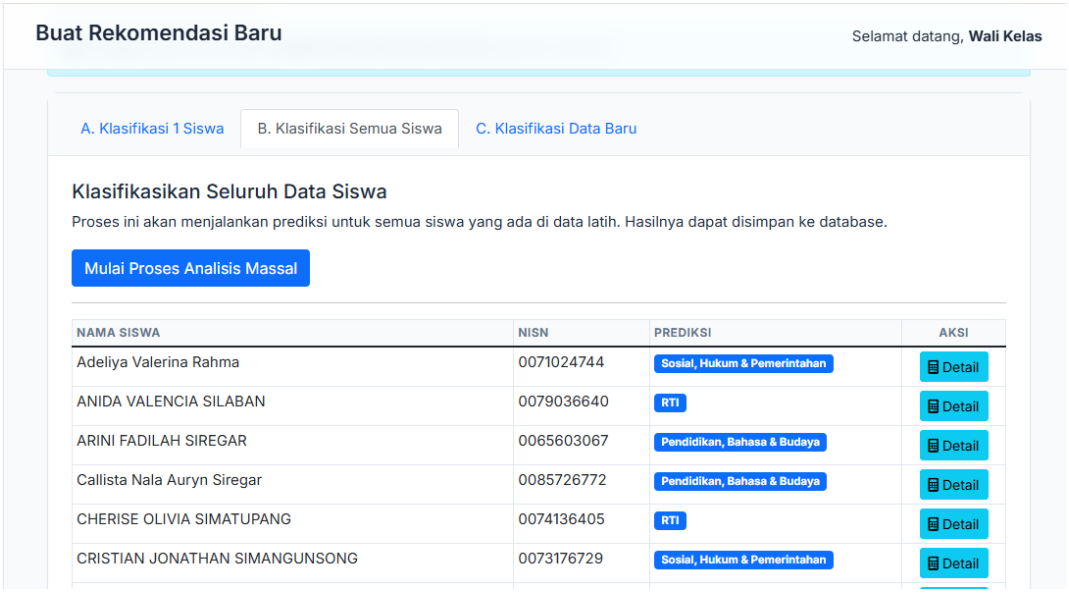
Tab ini digunakan untuk membuat rekomendasi bagi satu siswa spesifik yang datanya sudah ada di sistem. Wali kelas cukup memilih nama siswa, dan sistem akan langsung menampilkan laporan hasil rekomendasi lengkap dengan detail perhitungannya.



Gambar 4. 24 Klasifikasi 1 Siswa

b) Klasifikasi Semua Siswa

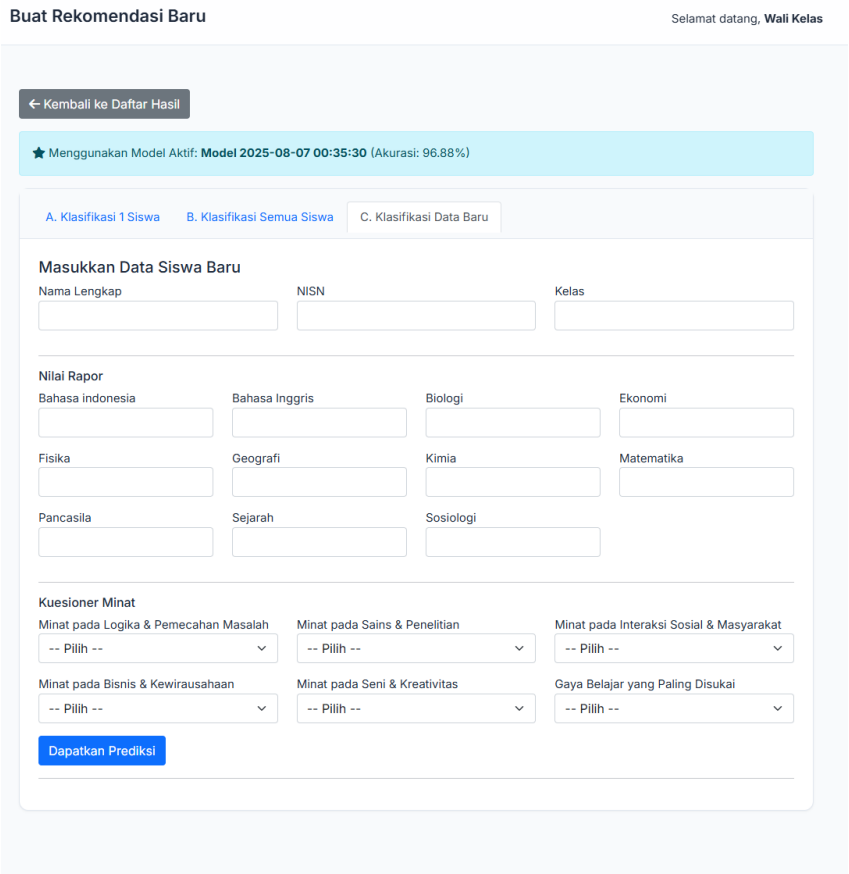
Fitur ini memungkinkan wali kelas untuk menjalankan prediksi secara massal untuk seluruh siswa yang terdaftar. Hasilnya ditampilkan dalam sebuah tabel ringkas, di mana setiap prediksi memiliki tombol "Detail" untuk melihat rincian perhitungannya.



Gambar 4. 25 Tampilan Klasifikasi Semua Siswa

c) Klasifikasi Data Baru

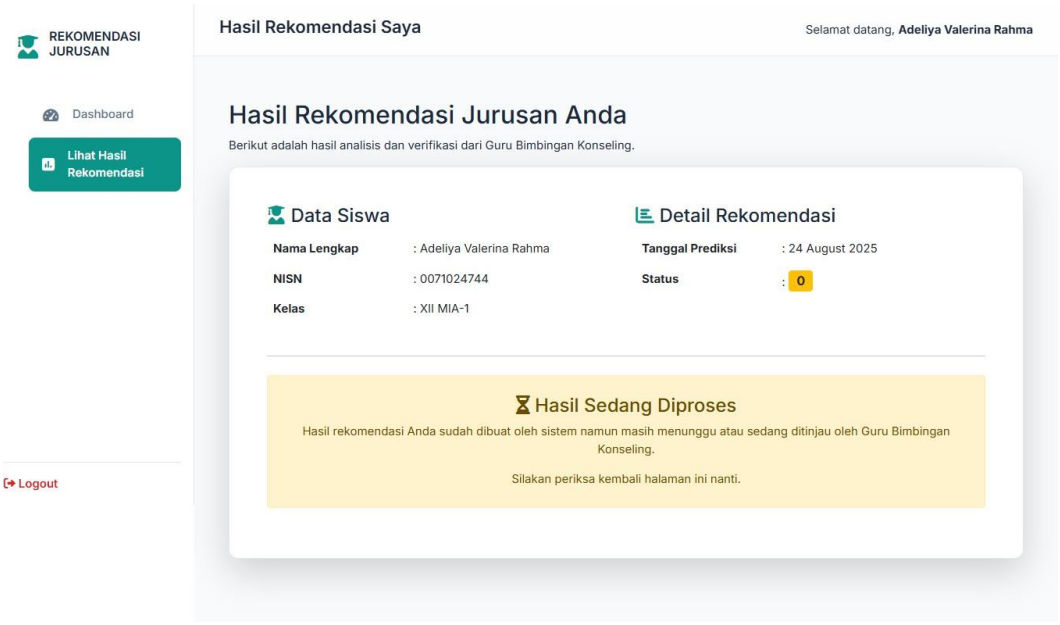
Tab ini berfungsi sebagai alat simulasi yang memungkinkan wali kelas untuk memasukkan data siswa yang sepenuhnya baru (nilai dan minat) dan langsung mendapatkan hasil prediksi tanpa harus menyimpan data siswa tersebut terlebih dahulu.



Gambar 4. 26 Klasifikasi Siswa Baru

4. Tampilan Halaman Lihat Hasil Rekomendasi (Siswa)

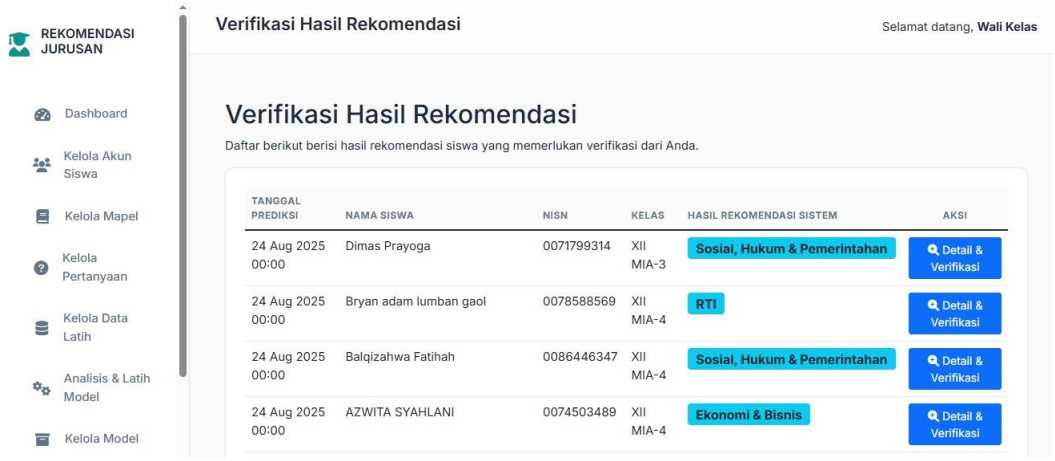
Di halaman ini, siswa melihat hasil akhir rekomendasi jurusan mereka. Jika hasil belum disetujui, siswa akan melihat status "Hasil Sedang Diproses"; jika sudah disetujui, halaman akan menampilkan jurusan yang direkomendasikan beserta catatan dari wali kelas.



Gambar 4. 27 Tampilan Hasil Rekomendasi (Siswa)

5. Tampilan Halaman Kelola Verifikasi Hasil

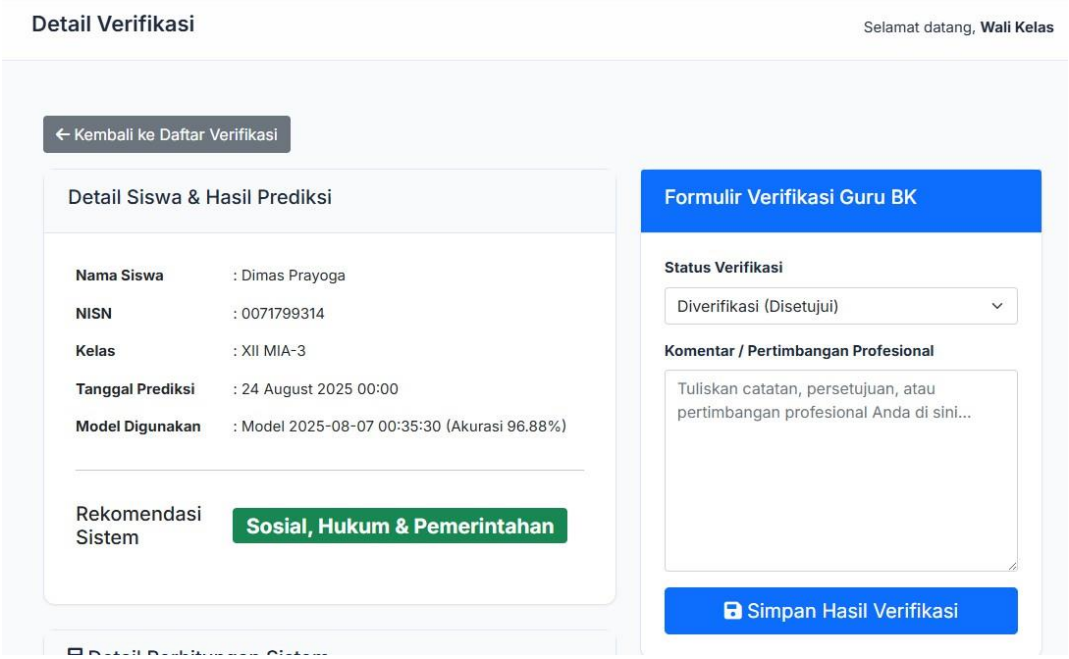
Halaman ini berfungsi sebagai antrean kerja bagi wali kelas untuk meninjau semua rekomendasi yang dihasilkan oleh sistem. Setiap rekomendasi yang masuk menunggu tindakan verifikasi dari wali kelas.



Gambar 4. 28 Tampilan Kelola Verifikasi Hasil

6. Tampilan Halaman Detail Verifikasi Hasil

Setelah memilih siswa dari daftar antrean, wali kelas diarahkan ke halaman ini untuk melakukan verifikasi akhir. Di sini, wali kelas dapat meninjau detail prediksi, melihat rincian perhitungan sistem, dan kemudian menyetujui atau menolak rekomendasi sambil memberikan komentar atau pertimbangan profesional mereka.



Gambar 4. 29 Tampilan Detail Verifikasi

4.3. Pengujian Program

Metode yang digunakan dalam pengujian adalah *Black Box Testing*, metode pengujian perangkat lunak yang berfokus pada fungsionalitas aplikasi dari sudut pandang pengguna, tanpa melihat atau mengetahui struktur kode internalnya. Metode ini menguji sistem dengan memberikan *input* dan memverifikasi *output* yang dihasilkan, untuk memastikan bahwa output tersebut sesuai dengan hasil yang diharapkan.

Untuk aplikasi ini, pengujian akan dilakukan untuk memvalidasi setiap fitur, mulai dari login, pengelolaan data oleh Wali Kelas, hingga proses siswa melihat hasil rekomendasinya. Hasil pengujian dapat dilihat pada tabel dibawah ini

Tabel 4. 10 Tabel Pengujian Program

No	Modul/Fitur yang Diuji	Skenario Pengujian	Langkah-langkah Pengujian	Hasil yang Diharapkan	Status
1	Autentikasi	Login Berhasil (Wali Kelas)	1) Buka halaman login.2) Masukkan email & password valid akun Wali Kelas.3) Klik **Login**.	Berhasil masuk dan diarahkan ke **Dashboard Wali Kelas**.	Berhasil
2	Autentikasi	Login Berhasil (Siswa)	1) Buka halaman login.2) Masukkan email & password valid akun Siswa.3) Klik **Login**.	Berhasil masuk dan diarahkan ke **Dashboard Siswa**.	Berhasil
3	Autentikasi	Login Gagal (Password Salah)	1) Buka halaman login.2) Masukkan email valid & password salah.3) Klik **Login**.	Tetap di halaman login dan muncul **notifikasi error**.	Sesuai

4	Autentikasi	Logout Pengguna	1) Login sebagai user apa pun.2) Klik tombol **Logout**.	Sesi berakhir dan diarahkan kembali ke **halaman login**.	Berhasil
5	Manajemen Data	Menambah Akun Siswa Baru	1) Login sebagai Wali Kelas.2) Buka **Kelola Akun Siswa**.3) Klik **Tambah Siswa Manual**.4) Isi form → **Simpan**.	Data siswa tersimpan dan muncul di **tabel siswa**.	Berhasil
6	Manajemen Data	Menghapus Akun Siswa	1) Login sebagai Wali Kelas.2) Buka **Kelola Akun Siswa**.3) Klik ikon **Hapus** pada salah satu siswa.4) Konfirmasi.	Data siswa terhapus dari **tabel siswa**.	Berhasil

7	Analisis & Model	Melatih Model Baru	1) Login sebagai Wali Kelas.2) Buka **Analisis & Latih Model**.3) Klik **Mulai Analisis Data**.	Proses analisis berjalan dan menampilkan **Hasil Analisis** (akurasi & penjelasan).	Berhasil
8	Analisis & Model	Menyimpan & Mengaktifkan Model	1) Setelah analisis selesai, klik **Simpan & Aktifkan Model Ini**.	Dialihkan ke **Kelola Model**; model baru muncul dengan status **Aktif**.	Berhasil
9	Rekomendasi	Membuat Rekomendasi 1 Siswa	1) Login Wali Kelas → **Hasil Rekomendasi → Buat Rekomendasi Baru**.2) Tab **Klasifikasi 1 Siswa**: pilih siswa → klik **Proses**.	Laporan rekomendasi tampil (jurusan prediksi + detail).	Berhasil

10	Rekomenda si	Menyimpan Hasil Rekomenda s i (1 Siswa)	1) Setelah laporan tampil, klik **Simpan Hasil**.	Hasil tersimpan dan muncul di **Daftar Hasil Rekomendasi** serta **Verifikasi Hasil**.	Berhasi l
11	Verifikasi	Menyetujui Rekomendas i	1) Login Wali Kelas → **Verifikasi Hasil**.2) Klik **Detail & Verifikasi** pada salah satu data.3) Pilih **Diverifikasi (Disetujui)**, isi komentar → **Simpan**.	Entri hilang dari antrean verifikasi; status pada data menjadi **Diverifikasi**.	Berhasil
12	Tampilan Siswa	Melihat Hasil (Belum Diverifikasi)	1) Login sebagai siswa dengan hasil **Belum Diverifikasi**. 2) Buka **Lihat Hasil Rekomendas i**.	Halaman menampilkan **status Belum Diverifikasi** / pesan **"Hasil sedang diproses"**.	Sesuai

13	Tampilan Siswa	Melihat Hasil (Sudah Diverifikasi)	1) Login sebagai siswa dengan hasil **Diverifikasi**. 2) Buka **Lihat Hasil Rekomendasi**	Tampil **jurusan direkomendasikan** dan **komentar wali kelas**.	Sesuai

4.4. Kelebihan dan Kelemahan Program

Analisis ini bertujuan untuk mengevaluasi secara fungsional dan teknis aplikasi "Sistem Rekomendasi Jurusan". Sistem ini dirancang sebagai alat pendukung keputusan bagi wali kelas memberikan rekomendasi jurusan kepada siswa berdasarkan data akademis dan minat, dengan alur kerja yang melibatkan verifikasi oleh tenaga ahli.

a. Kelebihan Program

Adapun kelebihan dari program ini adalah sebagai berikut :

1) Sistem tidak hanya memberikan hasil, tetapi juga menyajikan laporan detail mengenai proses perhitungan *Information Gain* dan Naive Bayes, sehingga proses pengambilan keputusan oleh AI menjadi transparan dan dapat dipahami oleh pengguna.
2) Adanya fitur verifikasi wajib oleh Guru BK memastikan bahwa hasil akhir tidak sepenuhnya bergantung pada mesin dan selalu divalidasi oleh keahlian profesional, menjadikannya alat bantu yang andal.
3) Aplikasi menyediakan fitur administratif yang komprehensif, mulai dari manajemen data siswa, mata pelajaran, hingga beberapa fitur yang sangat praktis untuk penggunaan di sekolah.

b. Kelemahan Program

Adapun kelemahan dari program ini adalah sebagai berikut :

1) Kecepatan analisis data bergantung pada spesifikasi komputer pengguna, sehingga bisa berjalan lambat pada perangkat lama.

2) Sistem belum terlalu efisien untuk sekolah dengan ribuan siswa karena harus memuat seluruh data sekaligus ke browser.
3) Kurangnya validasi di sisi *server* membuka celah bagi data yang tidak valid atau berbahaya untuk masuk ke sistem.

c. Kelebihan Program

Adapun kelebihan dari program ini adalah sebagai berikut :

1) Sistem tidak hanya memberikan hasil, tetapi juga menyajikan laporan detail mengenai proses perhitungan *Information Gain* dan Naive Bayes, sehingga proses pengambilan keputusan oleh AI menjadi transparan dan dapat dipahami oleh pengguna.
2) Adanya fitur verifikasi wajib oleh Guru BK memastikan bahwa hasil akhir tidak sepenuhnya bergantung pada mesin dan selalu divalidasi oleh keahlian profesional, menjadikannya alat bantu yang andal.
3) Aplikasi menyediakan fitur administratif yang komprehensif, mulai dari manajemen data siswa, mata pelajaran, hingga beberapa fitur yang sangat praktis untuk penggunaan di sekolah.

d. Kelemahan Program

Adapun kelemahan dari program ini adalah sebagai berikut :

1) Kecepatan analisis data bergantung pada spesifikasi komputer pengguna, sehingga bisa berjalan lambat pada perangkat lama.
2) Sistem belum terlalu efisien untuk sekolah dengan ribuan siswa karena harus memuat seluruh data sekaligus ke browser.
3) Kurangnya validasi di sisi *server* membuka celah bagi data yang tidak valid atau berbahaya untuk masuk ke sistem.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan proses perancangan, pengembangan, dan pengujian yang telah dilakukan terhadap Sistem Penentuan Jurusan, maka dapat ditarik kesimpulan sebagai berikut:

1. Sistem berbasis web berhasil dirancang dan dikembangkan dengan menggunakan pendekatan *hybrid* algoritma C4.5 dan Naïve Bayes yang mana C45 digunakan pada tahap seleksi fitur untuk menentukan atribut (nilai dan minat) yang paling berpengaruh dalam menentukan jurusan, serta Algoritma Naïve Bayes digunakan sebagai model klasifikasi utama untuk memprediksi jurusan berdasarkan fitur-fitur yang telah terpilih tersebut. Secara fungsional, sistem ini berhasil diwujudkan dalam sebuah aplikasi *prototype* yang memiliki alur kerja lengkap, mulai dari manajemen data siswa, proses analisis dan pelatihan model yang transparan, hingga verifikasi akhir oleh Wali Kelas. Perlu dicatat, implementasi akhir dari analisis data dan penggabungan metode dilakukan di sisi klien menggunakan JavaScript, yang memberikan pendekatan unik dan ringkas pada arsitektur sistem.
2. Berdasarkan hasil pengujian pada tab "Evaluasi Model", model *hybrid* yang dikembangkan terbukti efektif dalam memberikan rekomendasi jurusan. Sistem mampu mengukur dan menampilkan metrik kinerja secara kuantitatif, di antaranya :
    a) Akurasi: Dalam salah satu sesi pengujian, model berhasil mencapai akurasi sebesar 96.88%, yang menunjukkan tingkat ketepatan prediksi yang sangat tinggi.
    b) *Precision, Recall, dan F1-Score*: Sistem juga berhasil menampilkan nilai *precision*, *recall*, dan *F1-score* untuk setiap kelas (jurusan), memberikan gambaran detail mengenai seberapa baik model mengenali masing-masing jurusan.

Dengan demikian, dapat disimpulkan bahwa kinerja model sangat baik dan memenuhi tujuan penelitian untuk menghasilkan rekomendasi yang akurat.

5.2 Saran

Meskipun sistem telah berhasil dikembangkan sesuai dengan rumusan masalah, terdapat beberapa area yang dapat ditingkatkan di masa depan tanpa melanggar batasan masalah yang ada. Berikut adalah saran untuk pengembangan lanjutan:

1. Saat ini, sistem masih menyimpan *password* dalam bentuk teks biasa. Disarankan untuk mengimplementasikan fungsi *hashing* (misalnya password_hash() di PHP) pada fitur manajemen akun siswa. Peningkatan ini akan secara signifikan meningkatkan keamanan data pengguna jika terjadi kebocoran *database*, sejalan dengan tujuan pengujian fungsional dasar. Untuk meningkatkan akurasi model di masa depan, disarankan untuk menambahkan fitur di mana siswa dapat memberikan umpan balik (setuju/tidak setuju) terhadap hasil rekomendasi yang mereka terima. Data umpan balik ini nantinya dapat dikumpulkan dan digunakan sebagai variabel tambahan untuk melatih ulang model agar menjadi lebih akurat dan personal, tanpa mengubah metode inti *hybrid* C4.5 dan Naïve Bayes.
2. Untuk membantu Wali Kelas dalam proses verifikasi, disarankan untuk menambahkan fitur visualisasi data pada halaman detail verifikasi. Visualisasi ini dapat berupa diagram batang sederhana yang menampilkan nilai rapor dan skor minat siswa yang sedang ditinjau. Hal ini akan memberikan konteks yang lebih kaya bagi Wali Kelas saat mengevaluasi hasil rekomendasi dari sistem, sesuai dengan batasan pengolahan data nilai dan minat.

DAFTAR PUSTAKA

Alharbi, A. A., & Allohibi, J. (2024). A new hybrid classification algorithm for predicting student performance. *AIMS Mathematics*, *9*(7), 18308–18323. https://doi.org/10.3934/math.2024893

Ari Nurjani, M., Mulyadi, I., & Taryanto, A. (2021). Sistem Informasi Persediaan Barang Menggunakan Framework Bootstrap dan Bahasa Pemrograman PHP MYSQL. *CBIS JOURNAL*, *04*(02), 50–57. http://ejournal.upbatam.ac.id/index.php/cbis

Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: from efficient prediction to responsible AI. In *Frontiers in Artificial Intelligence* (Vol. 6). Frontiers Media SA. https://doi.org/10.3389/frai.2023.1124553

Hidayat, A., Yani, A., Rusidi, & Saadulloh. (2019). MEMBANGUN WEBSITE SMA PGRI GUNUNG RAYA RANAU MENGGUNAKAN PHP DAN MYSQL. *Jurnal*

Teknik Informatika Mahakarya, *2*(2), 41–52.

khoerul ummah. (2022). Sistem Informasi E-Learning Di SMA Negeri 1 Rantau Selatan Berbasis Web. *Journal of Student Development Informatics Management (JoSDIM)*, *1*(8.5.2017), 2003–2005.

Mustofa, K., & Kurniawan, D. (2024). *KESESUAIAN PEMILIHAN JURUSAN SISWA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA SMK PLUS AL MUSYARROFAH* (Vol. 06, Issue 1).

Naldo, N., Sidratul Munti, N. Y., & Adeswastoto, H. (2021). Design and Build a Website-Based Mechanic Information System in Bangkinang City (Analysis

System). *Journal of Engineering Science and Technology Management | ISSN*, *1*(2), 2828–7886.

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada

Twitter. In *Jurnal Sains Komputer & Informatika (J-SAKTI* (Vol. 5, Issue 2).

Rasekh, N., & Bavil, D. N. (n.d.). Improvement accuracy for C4.5 decision tree algorithm. *Int. J. Nonlinear Anal. Appl. In Press*, 2008–6822. https://doi.org/10.22075/ijnaa.2024.33099.4925

Rumbaugh, James., Jacobson, Ivar., & Booch, Grady. (2021). *The unified modeling language reference manual*. Addison-Wesley.

Saputra, D., Irmayani, W., Purwaningtias, D., & Sidauruk, J. (2021). A Comparative Analysis of C4.5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction. *International Journal of Advances in Data and Information Systems*,

2(2). https://doi.org/10.25008/ijadis.v2i2.1221

Setiawan, S. (2020). *Membicarakan Precision, Recall, dan F1-Score*. Medium.Com. https://stevkarta.medium.com/membicarakan-precision-recall-dan-f1-score- e96d81910354

Unesa. (2025). *Sistem Pendukung Keputusan: Pengertian, Komponen, dan Manfaat dalam Pengambilan Keputusan*. https://si.ft.unesa.ac.id/post/sistem-pendukung- keputusan-pengertian-komponen-dan-manfaat-dalam-pengambilan-keputusan?utm_source=chatgpt.com

Yudhi Putra, M., & Ismiyana Putri, D. (2023). *Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI* (Vol. 16, Issue 2).

Yusuf, P. M. (2022). Design of Semi Automatic Announcement System In The Public Address System Part of The Main Branch Office Soekarno-Hatta International Airport Angkasa Pura II, Inc. In *Journal of Electrical, Electronic, Information, and Communication Technology (JEEICT)* (Vol. 4, Issue 2).