

Analisis Penentuan Kelulusan Sesuai Prestasi Siswa Menggunakan Algoritma K-Nearest Neighbor

Raynaldy¹, Erfan Hasmin^{2*}, Nurul Aini²

¹Teknik Informatika, Universita Dipa Makassar

²Universitas Dipa Makassar, Jln. Perintis Kemerdekaan KM 9 Kota Makassar

e-mail: ¹raynaldyraynaldy@gmail.com, ^{2*}erfan.hasmin@undipa.ac.id, ³nurulaini.m11@undipa.ac.id

Abstrak

Penelitian ini mengembangkan Modul Prediksi Akademik yang terintegrasi pada sistem e-Rapor berbasis web menggunakan algoritma K-Nearest Neighbor (KNN) untuk mengidentifikasi potensi kelulusan akademik siswa berdasarkan pola nilai rapor. Hasil prediksi tidak digunakan sebagai keputusan akhir kelulusan, melainkan sebagai informasi pendukung dan sistem peringatan dini akademik dengan dua kategori, yaitu Berpotensi Lulus dan Berisiko Tidak Lulus. Data penelitian berasal dari nilai akhir siswa SMA Kristen Elim Makassar tahun akademik 2025/2026 semester Ganjil. Setelah melalui proses seleksi, pencocokan NIS, penghapusan duplikasi, dan validasi nilai, diperoleh 283 data siswa valid, terdiri atas 242 siswa Berpotensi Lulus dan 41 siswa Berisiko Tidak Lulus berdasarkan label acuan akademik. Data dinormalisasi menggunakan Min-Max Normalization, kemudian diklasifikasikan menggunakan Euclidean Distance dengan pengujian $k=3$, $k=5$, dan $k=7$ pada beberapa skenario pembagian data. Hasil terbaik diperoleh pada pembagian data 80:20 dengan $k=3$, menghasilkan akurasi 87,72%, precision 87,23%, recall 97,62%, dan F1-score 92,13%. Hasil tersebut menunjukkan bahwa integrasi KNN pada e-Rapor mampu menyediakan informasi prediktif yang mendukung pemantauan kondisi akademik siswa secara lebih terstruktur, tanpa menggantikan kewenangan sekolah dalam menentukan kelulusan akhir.

Kata kunci: E-Rapor, K-Nearest Neighbor, Prediksi Akademik, Data Mining, Sistem Peringatan Dini.

Abstract

This study develops an Academic Prediction Module integrated into a web-based e-report system using the K-Nearest Neighbor (KNN) algorithm to identify students' academic graduation potential based on report-card score patterns. The prediction results are not used as final graduation decisions, but as supporting information and an academic early warning system with two categories: Potentially Passing and At Risk of Not Passing. The research data were obtained from the final grades of students at SMA Kristen Elim Makassar in the odd semester of the 2025/2026 academic year. After data selection, Student Identification Number (NIS) matching, duplicate removal, and grade validation, 283 valid student records were obtained, consisting of 242 students in the Potentially Passing class and 41 students in the At Risk of Not Passing class based on the academic reference labels. The data were normalized using Min-Max Normalization and classified using Euclidean Distance with $k=3$, $k=5$, and $k=7$ under several train-test split scenarios. The best performance was achieved using an 80:20 split with $k=3$, producing an accuracy of 87.72%, precision of 87.23%, recall of 97.62%, and F1-score of 92.13%. These results indicate that integrating KNN into the e-report system can provide predictive information to support more structured monitoring of students' academic conditions without replacing the school's authority in determining final graduation status.

Keywords: E-Report, K-Nearest Neighbor, Academic Prediction, Data Mining, Academic Early Warning System.

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong lembaga pendidikan untuk melakukan digitalisasi dalam pengelolaan data akademik. Penerapan sistem berbasis web memungkinkan proses penginputan, pengolahan, penyimpanan, dan penyajian nilai dilakukan secara lebih terstruktur dibandingkan pengelolaan secara manual [1], [2]. Salah satu bentuk penerapan tersebut adalah sistem e-Rapor yang digunakan untuk membantu guru dan pihak sekolah dalam mengelola hasil belajar siswa serta menyediakan

informasi akademik secara lebih cepat dan terorganisasi [3], [4]. Meskipun demikian, pemanfaatan data pada sistem e-Rapor masih banyak berorientasi pada pencatatan dan pelaporan hasil belajar, sedangkan data akademik yang tersimpan memiliki potensi untuk dikembangkan menjadi sumber informasi analitik yang dapat mendukung pemantauan kondisi akademik siswa.

Sistem e-Rapor berbasis web memberikan kemudahan dalam pengelolaan nilai, penyusunan laporan, dan akses informasi hasil belajar bagi pihak sekolah maupun orang tua [1], [5]. Efektivitas penggunaan e-Rapor juga menunjukkan bahwa sistem digital dapat membantu meningkatkan efisiensi proses pengelolaan dan pelaporan akademik [2]. Namun, keberadaan data nilai yang tersimpan secara digital belum secara langsung menghasilkan informasi prediktif. Oleh karena itu, diperlukan pengembangan yang tidak hanya memanfaatkan e-Rapor sebagai media administrasi, tetapi juga mampu mengolah data nilai menjadi informasi pendukung untuk mengenali kondisi akademik siswa secara lebih dini.

Pada SMA Kristen Elim Makassar, sistem e-Rapor digunakan untuk mengelola data siswa, kelas, mata pelajaran, proses penilaian, nilai akhir, dan laporan hasil belajar. Data tersebut merepresentasikan capaian akademik siswa pada setiap periode pembelajaran. Apabila data hanya digunakan untuk menghasilkan laporan, maka pemanfaatannya masih terbatas pada dokumentasi hasil belajar. Pengembangan fungsi analitik diperlukan agar data nilai dapat memberikan informasi tambahan mengenai siswa yang memiliki kecenderungan akademik baik maupun siswa yang memerlukan perhatian lebih awal. Dalam penelitian ini, informasi tersebut disajikan dalam dua kategori, yaitu Berpotensi Lulus dan Berisiko Tidak Lulus, sebagai bentuk dukungan pemantauan akademik dan bukan sebagai keputusan akhir kelulusan.

Salah satu pendekatan yang dapat digunakan untuk melakukan klasifikasi terhadap data akademik adalah K-Nearest Neighbor (KNN). KNN merupakan metode klasifikasi berbasis kedekatan yang menentukan kelas suatu data berdasarkan mayoritas kelas dari sejumlah tetangga terdekat [6]. Penerapan KNN pada bidang pendidikan telah digunakan untuk memprediksi kelulusan dan mengklasifikasikan prestasi akademik berdasarkan karakteristik data siswa [7], [8], [9]. Dalam penelitian ini, KNN diterapkan pada pola nilai rapor mata pelajaran yang valid. Sebelum proses klasifikasi, data dinormalisasi menggunakan Min-Max Normalization, kemudian kedekatan antar data dihitung menggunakan Euclidean Distance. Pendekatan tersebut memungkinkan proses klasifikasi mempertimbangkan pola beberapa atribut nilai, bukan hanya bergantung pada satu nilai rekapitulasi.

Sejumlah penelitian sebelumnya telah menunjukkan penerapan KNN dalam analisis akademik. Penelitian [7] menerapkan KNN untuk prediksi kelulusan siswa, sedangkan penelitian [8] menggunakan KNN untuk memprediksi tingkat kelulusan peserta pendidikan profesi guru. Penelitian [9] juga menunjukkan penerapan KNN untuk mengklasifikasikan prestasi akademik siswa berdasarkan nilai rapor dan kedisiplinan. Pada sisi lain, penelitian mengenai e-Rapor lebih banyak membahas pengembangan, efektivitas, dan fungsi sistem dalam pengelolaan serta pelaporan nilai [1], [2], [3], [4], [5]. Berdasarkan referensi tersebut, masih terdapat ruang pengembangan untuk mengintegrasikan proses klasifikasi akademik secara langsung ke dalam sistem e-Rapor yang digunakan dalam pengelolaan nilai, sehingga hasil analisis dapat disajikan bersama data akademik dalam satu sistem.

Kebaruan penelitian ini terletak pada pengembangan Modul Prediksi Akademik berbasis KNN yang terintegrasi dengan sistem e-Rapor berbasis web. Modul tersebut memanfaatkan nilai rapor yang telah tersimpan dalam sistem untuk menghasilkan informasi prediktif dan menyajikannya melalui dashboard pada tingkat sekolah, kelas, dan individu siswa. Selain menampilkan kategori Berpotensi Lulus dan Berisiko Tidak Lulus, sistem juga menyediakan ringkasan prediksi, rekapitulasi per kelas, detail prediksi siswa, statistik kelas, tingkat keyakinan hasil klasifikasi, evaluasi model, serta riwayat prediksi berdasarkan periode akademik. Dengan demikian, kontribusi penelitian tidak hanya terletak pada penerapan algoritma KNN, tetapi juga pada integrasinya dengan sistem e-Rapor sebagai alat bantu analisis dan sistem peringatan dini akademik.

Berdasarkan permasalahan dan kesenjangan tersebut, penelitian ini bertujuan mengembangkan dan mengevaluasi Modul Prediksi Akademik pada sistem e-Rapor berbasis web menggunakan algoritma K-Nearest Neighbor. Model digunakan untuk mengklasifikasikan siswa ke dalam kategori Berpotensi Lulus dan Berisiko Tidak Lulus berdasarkan pola nilai akademik. Kinerja model dievaluasi menggunakan beberapa skenario pembagian data dan nilai k dengan metrik accuracy, precision, recall, F1-score, dan confusion matrix. Hasil klasifikasi digunakan sebagai informasi pendukung bagi guru dan pihak sekolah dalam melakukan pemantauan serta identifikasi dini kondisi akademik siswa. Penentuan kelulusan resmi tetap harus mempertimbangkan proses evaluasi pendidikan dan pencapaian kompetensi peserta didik secara lebih menyeluruh [10].

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk menganalisis potensi kelulusan akademik siswa berdasarkan data prestasi belajar. Algoritma yang digunakan adalah K-Nearest Neighbor (KNN), yaitu metode klasifikasi yang bekerja dengan membandingkan kedekatan data uji terhadap data latih berdasarkan jarak antar atribut. Dalam penelitian ini, KNN digunakan untuk mengelompokkan siswa ke dalam dua kategori, yaitu Berpotensi Lulus dan Berisiko Tidak Lulus, berdasarkan pola nilai akademik yang tersimpan pada sistem e-Rapor. Hasil klasifikasi tersebut digunakan sebagai informasi pendukung bagi sekolah, bukan sebagai keputusan akhir kelulusan siswa.

2.1. Jenis dan Sumber Data Penelitian

Jenis data yang digunakan dalam penelitian ini adalah data kuantitatif berupa nilai akademik siswa yang diperoleh dari sistem e-Rapor SMA Kristen Elim Makassar. Data yang dianalisis merupakan nilai akhir siswa pada tahun akademik 2025/2026 semester Ganjil. Pemilihan data semester Ganjil dilakukan karena data tersebut dapat digunakan sebagai gambaran awal untuk menganalisis potensi kelulusan akademik siswa pada periode berjalan.

Sumber data berasal dari database sistem e-Rapor yang memuat data siswa, data kelas, data tahun akademik, nilai akhir, serta detail nilai rapor setiap mata pelajaran. Setelah melalui proses seleksi berdasarkan periode akademik, pencocokan Nomor Induk Siswa (NIS), pembersihan data, penghapusan duplikasi, dan validasi nilai, diperoleh 283 data siswa yang digunakan sebagai dataset penelitian.

Atribut yang digunakan dalam proses klasifikasi berasal dari nilai rapor setiap mata pelajaran yang valid. Nilai yang kosong, tidak berbentuk angka, atau mata pelajaran yang tidak berlaku bagi siswa tertentu tidak digunakan dalam perhitungan. Atribut administratif seperti alamat, nama orang tua, dan informasi wali tidak digunakan karena tidak berhubungan langsung dengan capaian akademik siswa. Label klasifikasi terdiri atas dua kelas, yaitu Berpotensi Lulus dan Berisiko Tidak Lulus. Siswa dikategorikan Berpotensi Lulus apabila rata-rata nilai mata pelajaran minimal 80 dan tidak memiliki nilai mata pelajaran valid di bawah 80. Sebaliknya, siswa dikategorikan Berisiko Tidak Lulus apabila rata-rata nilai mata pelajaran kurang dari 80 atau terdapat nilai mata pelajaran valid di bawah 80.

2.2. Tahapan Penelitian

Tahapan penelitian dimulai dengan pengumpulan data nilai akhir siswa dari sistem e-Rapor SMA Kristen Elim Makassar pada tahun akademik 2025/2026 semester Ganjil. Data tersebut digunakan sebagai dasar untuk mengklasifikasikan potensi kelulusan akademik siswa menggunakan algoritma K-Nearest Neighbor (KNN).

Tahap berikutnya adalah preprocessing data. Pada tahap ini dilakukan seleksi data berdasarkan tahun akademik dan semester, pencocokan NIS dengan data siswa, penghapusan data duplikat, serta validasi nilai rapor setiap mata pelajaran. Nilai yang berada pada rentang 0 sampai 100 digunakan sebagai atribut, sedangkan nilai kosong, tanda “-”, atau mata pelajaran yang tidak berlaku tidak dimasukkan dalam perhitungan.

Setelah data siap digunakan, sistem membentuk dataset KNN berdasarkan nilai rapor mata pelajaran yang valid. Dataset kemudian diuji menggunakan beberapa skenario pembagian data, yaitu 70:30, 75:25, 80:20, 85:15, dan 90:10. Pada setiap skenario, sistem menguji nilai $k = 3$, $k = 5$, dan $k = 7$ untuk melihat performa model pada kombinasi data latih dan data uji yang berbeda.

Tahap selanjutnya adalah proses klasifikasi menggunakan KNN. Data terlebih dahulu dinormalisasi menggunakan Min-Max normalization, kemudian jarak antara data uji dan data latih dihitung menggunakan Euclidean Distance. Hasil klasifikasi ditentukan berdasarkan mayoritas label dari sejumlah tetangga terdekat. Tahap terakhir adalah evaluasi model menggunakan accuracy, precision, recall, F1-score, dan confusion matrix.

2.3. Metode K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan algoritma klasifikasi yang menentukan kelas suatu data berdasarkan kedekatannya dengan sejumlah data lain yang telah memiliki label. Prinsip utama KNN adalah bahwa data dengan karakteristik yang mirip cenderung berada pada kelas yang sama. Dalam penelitian ini, kemiripan antar siswa dihitung berdasarkan pola nilai rapor setiap mata pelajaran yang valid.

Sebelum perhitungan jarak dilakukan, data dinormalisasi menggunakan Min-Max normalization. Normalisasi ini digunakan agar setiap atribut nilai berada pada skala yang seimbang, sehingga tidak ada atribut yang memiliki pengaruh terlalu besar hanya karena perbedaan rentang nilai. Rumus Min-Max normalization adalah sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan:

X' adalah nilai hasil normalisasi, X adalah nilai awal, X_{min} adalah nilai minimum pada atribut, dan X_{max} adalah nilai maksimum pada atribut.

Setelah proses normalisasi, jarak antara data uji dan data latih dihitung menggunakan Euclidean Distance. Rumus Euclidean Distance adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$ adalah jarak antara data uji dan data latih, x_i adalah nilai atribut pada data uji, dan y_i adalah nilai atribut pada data latih.

Nilai k yang digunakan dalam penelitian ini adalah $k = 3$, $k = 5$, dan $k = 7$. Penggunaan nilai k ganjil bertujuan untuk mengurangi kemungkinan hasil klasifikasi yang seimbang antara dua kelas. Pada setiap nilai k , sistem mengurutkan jarak dari yang terkecil, mengambil sejumlah tetangga terdekat, kemudian menentukan kelas berdasarkan label yang paling dominan. Hasil klasifikasi ditampilkan dalam dua kategori, yaitu Berpotensi Lulus dan Berisiko Tidak Lulus.

Sistem juga dapat menampilkan tingkat keyakinan voting KNN berdasarkan proporsi kelas mayoritas:

$$Confidencex = \frac{N_{mayoritas}}{k} \times 100\%$$

Nilai tersebut merupakan tingkat kesepakatan tetangga terdekat, bukan probabilitas pasti seorang siswa akan lulus.

2.4. Evaluasi Kinerja

Evaluasi kinerja dilakukan untuk mengetahui kemampuan algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan potensi kelulusan akademik siswa. Evaluasi tidak hanya menggunakan akurasi, tetapi juga precision, recall, F1-score, dan confusion matrix agar performa model dapat dianalisis secara lebih menyeluruh.

Akurasi digunakan untuk mengukur perbandingan antara jumlah prediksi yang benar dengan total data uji. Rumus akurasi adalah sebagai berikut:

$$Accuracy = \frac{\text{Jumlah prediksi benar}}{\text{Total data}} \times 100\%$$

Precision digunakan untuk mengukur ketepatan model dalam menghasilkan prediksi pada kelas Berpotensi Lulus. Rumus precision adalah sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

Recall digunakan untuk mengukur kemampuan model dalam mengenali data yang termasuk kelas Berpotensi Lulus. Rumus recall adalah sebagai berikut:

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

F1-score digunakan untuk melihat keseimbangan antara precision dan recall. Rumus F1-score adalah sebagai berikut:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Selain itu, confusion matrix digunakan untuk melihat distribusi hasil klasifikasi dalam bentuk TP, TN, FP, dan FN. TP menunjukkan data Berpotensi Lulus yang diprediksi benar sebagai Berpotensi Lulus. TN menunjukkan data Berisiko Tidak Lulus yang diprediksi benar sebagai Berisiko Tidak Lulus. FP menunjukkan data Berisiko Tidak Lulus yang diprediksi sebagai Berpotensi Lulus. FN menunjukkan data Berpotensi Lulus yang diprediksi sebagai Berisiko Tidak Lulus. Evaluasi dilakukan pada beberapa skenario pembagian data dan nilai k agar hasil pengujian tidak bergantung pada satu konfigurasi data saja.

2.5. Perancangan Sistem

Sistem yang dikembangkan merupakan sistem e-Rapor berbasis web yang dilengkapi dengan modul Prediksi Akademik. Sistem ini dirancang untuk mengelola data akademik siswa, menyimpan nilai akhir dan detail nilai setiap mata pelajaran, memproses klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN), serta menampilkan hasil prediksi dalam bentuk dashboard.

Alur kerja sistem dimulai dari proses input nilai oleh guru mata pelajaran. Nilai tersebut diolah menjadi nilai rapor dan direkap pada fitur Nilai Akhir. Setelah disimpan, sistem mencatat rekap nilai siswa beserta detail nilai setiap mata pelajaran yang menjadi atribut utama dalam proses KNN. Data tersebut kemudian digunakan oleh modul Prediksi Akademik untuk melakukan klasifikasi potensi kelulusan

akademik siswa.

Hasil prediksi ditampilkan dalam dua kategori, yaitu Berpotensi Lulus dan Berisiko Tidak Lulus. Modul Prediksi Akademik juga menampilkan ringkasan prediksi seluruh sekolah, rekap prediksi per kelas, detail prediksi siswa, statistik per kelas, riwayat prediksi per semester, serta evaluasi model KNN. Dengan rancangan ini, sistem e-Rapor tidak hanya berfungsi sebagai media pencatatan nilai, tetapi juga sebagai alat bantu analisis akademik dan peringatan dini bagi pihak sekolah.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Sistem

Sistem yang dikembangkan dalam penelitian ini merupakan sistem e-Rapor berbasis web yang dirancang untuk mendukung pengelolaan data akademik siswa di SMA Kristen Elim Makassar. Sistem ini tidak hanya digunakan untuk mencatat dan menampilkan hasil belajar, tetapi juga dikembangkan agar data akademik yang tersimpan dapat dimanfaatkan sebagai dasar analisis potensi kelulusan siswa. Melalui penerapan algoritma K-Nearest Neighbor (KNN), nilai akademik yang telah diolah dalam sistem dapat digunakan untuk mengidentifikasi siswa yang Berpotensi Lulus dan siswa yang Berisiko Tidak Lulus secara akademik.

Secara fungsional, sistem e-Rapor mendukung beberapa proses utama, yaitu pengelolaan data siswa, pengelolaan data guru, pengaturan tahun akademik, input nilai, pengolahan nilai akhir, pencetakan rapor, dan prediksi akademik. Nilai akhir siswa tetap menjadi sumber utama dalam proses klasifikasi, sedangkan hasil prediksi ditampilkan melalui modul Prediksi Akademik. Dengan pemisahan ini, fitur Nilai Akhir berperan sebagai penyedia data akademik, sementara modul Prediksi Akademik berperan sebagai media analisis dan penyajian hasil klasifikasi.

Sistem ini menggunakan pembagian hak akses berdasarkan peran pengguna. Aktor yang terlibat meliputi admin, guru mata pelajaran, guru BK, wali kelas, dan wali murid. Admin bertanggung jawab mengelola data utama sistem, guru mata pelajaran mengelola penilaian akademik, guru BK mengelola data absensi dan sikap siswa, wali kelas memantau nilai akhir serta mencetak laporan akademik, sedangkan wali murid memperoleh akses untuk melihat informasi hasil belajar siswa. Pembagian peran tersebut membuat pengelolaan data lebih terkontrol karena setiap pengguna hanya dapat mengakses fitur sesuai dengan tanggung jawabnya.

Dalam konteks penerapan KNN, data yang digunakan berasal dari nilai akhir siswa tahun akademik 2025/2026 semester Ganjil. Data tersebut diproses melalui seleksi periode akademik, validasi nilai, dan preprocessing sebelum digunakan dalam klasifikasi. Hasil yang ditampilkan sistem berupa kategori Berpotensi Lulus dan Berisiko Tidak Lulus. Informasi ini berfungsi sebagai alat bantu analisis akademik dan peringatan dini bagi pihak sekolah, bukan sebagai keputusan akhir kelulusan siswa.

3.2. Use Case Diagram Sistem

Use case diagram digunakan untuk menggambarkan hubungan antara aktor dengan fungsi utama yang tersedia pada sistem e-Rapor berbasis web. Diagram ini menunjukkan bagaimana setiap pengguna berinteraksi dengan sistem sesuai dengan peran dan hak akses masing-masing. Pada sistem ini, seluruh aktor harus melalui proses login dan memilih peran sebelum dapat mengakses fitur yang tersedia. Proses login juga dilengkapi dengan verifikasi password dan pesan kesalahan apabila data yang dimasukkan tidak sesuai.

Berdasarkan use case diagram pada gambar 1, terdapat enam aktor utama yang terlibat dalam sistem, yaitu admin, operator data, guru mata pelajaran, guru BK, wali kelas, dan wali murid. Setiap aktor memiliki ruang lingkup akses yang berbeda agar pengelolaan data dapat dilakukan secara terkontrol sesuai tanggung jawab masing-masing pengguna.

Admin memiliki hak akses paling luas dalam sistem. Admin dapat mengelola data user, data siswa, data guru, data sekolah, data tambahan, mata pelajaran, daftar wali kelas dan guru mapel, nilai akhir, catatan wali kelas, serta modul Prediksi Akademik. Peran admin menjadi penting karena data yang dikelola menjadi dasar bagi proses akademik lainnya, termasuk pengolahan nilai dan klasifikasi potensi kelulusan akademik siswa.

Operator data berperan dalam membantu pengelolaan data utama yang berkaitan dengan administrasi akademik. Berdasarkan diagram, operator data dapat mengakses pengelolaan data siswa, data guru, data tambahan, serta daftar wali kelas dan guru mapel. Keberadaan operator data membantu proses pemeliharaan data agar informasi yang digunakan dalam sistem tetap lengkap dan mutakhir.

Guru mata pelajaran berperan dalam mengelola data pembelajaran dan penilaian akademik. Fitur yang dapat diakses meliputi tujuan pembelajaran, lingkup materi, asesmen formatif, asesmen sumatif,

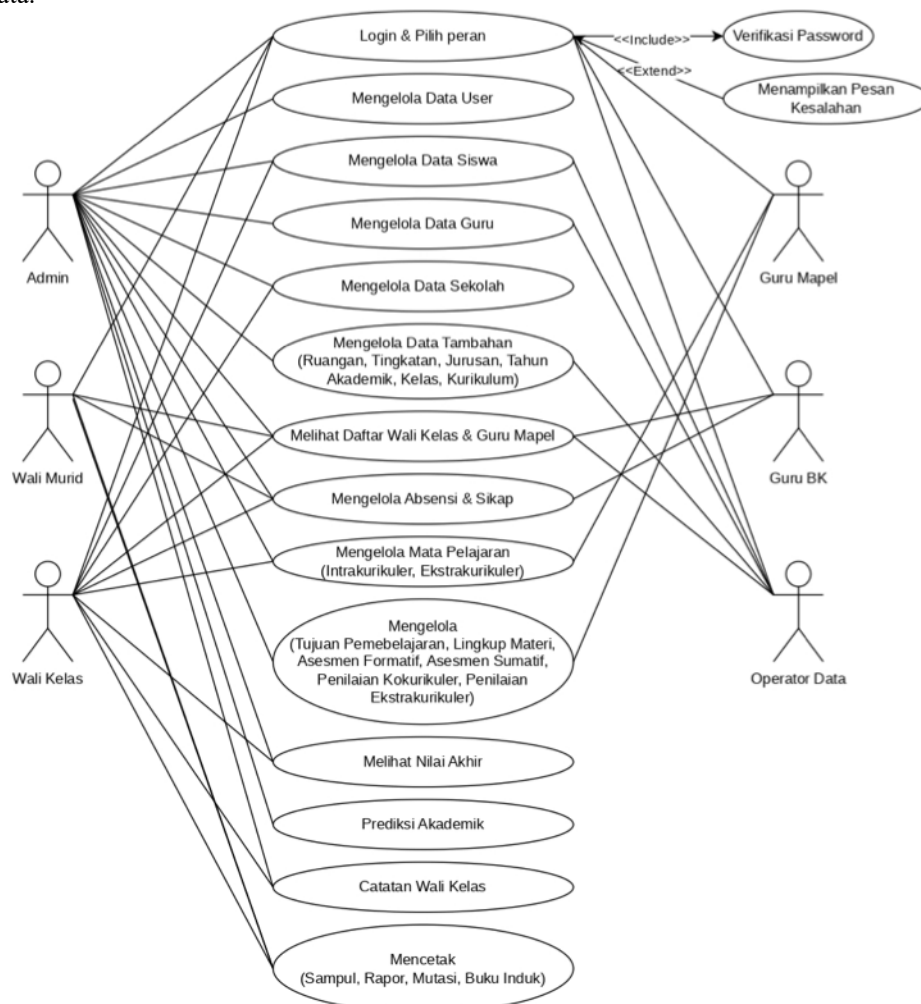
penilaian kokurikuler, dan penilaian ekstrakurikuler. Data penilaian yang diinput oleh guru mata pelajaran menjadi sumber utama dalam pembentukan nilai akhir siswa.

Guru BK memiliki akses terhadap data absensi dan sikap siswa. Data ini digunakan sebagai informasi pendukung dalam sistem e-Rapor. Meskipun proses KNN pada penelitian ini berfokus pada data nilai akademik, keberadaan data absensi dan sikap tetap penting karena dapat menjadi bahan pengembangan pada penelitian selanjutnya untuk memperluas variabel analisis akademik.

Wali kelas memiliki peran dalam memantau perkembangan akademik siswa, melihat nilai akhir, mencetak dokumen akademik, serta mengelola catatan wali kelas. Dokumen yang dapat dicetak meliputi sampul rapor, rapor, mutasi, dan buku induk. Selain itu, catatan wali kelas menjadi bagian penting dalam melengkapi informasi perkembangan siswa secara naratif.

Wali murid memperoleh akses untuk melihat informasi hasil belajar siswa sesuai hak akses yang diberikan. Dengan adanya akses ini, sistem e-Rapor tidak hanya digunakan oleh pihak internal sekolah, tetapi juga mendukung transparansi informasi akademik kepada orang tua atau wali siswa.

Secara keseluruhan, use case diagram menunjukkan bahwa sistem e-Rapor dirancang dengan pembagian peran yang jelas. Pengelolaan data, penginputan nilai, pencetakan rapor, catatan wali kelas, dan Prediksi Akademik ditempatkan sesuai kebutuhan masing-masing aktor. Pembagian akses tersebut membuat sistem lebih terstruktur dan mendukung pengelolaan informasi akademik secara lebih aman, terarah, dan berbasis data.



Gambar 1. Use Case Diagram Sistem E-Rapor Berbasis Web

3.3. Analisis Data Siswa dan Nilai Akademik

Data yang digunakan dalam penelitian ini berasal dari database sistem e-Rapor SMA Kristen Elim Makassar. Data tersebut mencakup data siswa, kelas, tahun akademik, nilai akhir, serta detail nilai rapor setiap

mata pelajaran. Data siswa digunakan sebagai identitas utama, sedangkan data nilai akademik digunakan sebagai dasar klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN).

Tidak seluruh data siswa digunakan sebagai atribut dalam proses klasifikasi. Atribut administratif seperti alamat, nama orang tua, dan informasi wali tidak digunakan karena tidak berkaitan langsung dengan capaian akademik. Penelitian ini hanya memanfaatkan data nilai yang merepresentasikan prestasi belajar siswa, yaitu nilai rapor dari setiap mata pelajaran yang valid.

Data nilai yang dianalisis merupakan nilai akhir siswa pada tahun akademik 2025/2026 semester Ganjil. Pemilihan semester Ganjil dilakukan karena data tersebut dapat memberikan gambaran awal mengenai kondisi akademik siswa pada periode berjalan. Nilai yang digunakan dalam proses KNN adalah nilai numerik pada rentang 0 sampai 100. Sementara itu, nilai kosong, tanda “-”, atau mata pelajaran yang tidak berlaku bagi siswa tertentu tidak dimasukkan dalam perhitungan agar tidak memengaruhi hasil klasifikasi.

3.4. Pembentukan Dataset Prediksi Akademik

Dataset prediksi akademik dibentuk dari nilai akhir siswa yang tersimpan pada sistem e-Rapor. Pada sistem yang dikembangkan, nilai akhir tidak hanya disimpan dalam bentuk rekap, tetapi juga dalam bentuk detail nilai rapor setiap mata pelajaran. Detail nilai inilah yang digunakan sebagai atribut utama dalam proses klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN). Dengan demikian, setiap siswa direpresentasikan sebagai satu data yang memuat pola nilai pada mata pelajaran yang valid.

Sebelum digunakan dalam proses klasifikasi, data melalui tahap preprocessing. Tahap ini mencakup seleksi data berdasarkan tahun akademik dan semester, pencocokan NIS dengan data siswa, penghapusan data duplikat, serta validasi nilai rapor. Nilai yang digunakan sebagai atribut adalah nilai numerik dengan rentang 0 sampai 100. Nilai kosong, tanda “-”, atau mata pelajaran yang tidak berlaku bagi siswa tertentu tidak digunakan dalam perhitungan agar tidak dianggap sebagai nilai rendah atau nilai nol.

Setelah tahap preprocessing, diperoleh 283 data siswa yang digunakan sebagai dataset penelitian. Label klasifikasi dibentuk berdasarkan aturan akademik yang diterapkan pada sistem. Siswa dikategorikan Berpotensi Lulus apabila rata-rata nilai mata pelajaran minimal 80 dan tidak memiliki nilai mata pelajaran valid di bawah 80. Sebaliknya, siswa dikategorikan Berisiko Tidak Lulus apabila rata-rata nilai mata pelajaran kurang dari 80 atau terdapat nilai mata pelajaran valid di bawah 80.

Pembentukan dataset dengan menggunakan nilai rapor setiap mata pelajaran bertujuan agar proses KNN tidak hanya bergantung pada satu nilai rekapitulasi. Melalui pendekatan ini, sistem dapat membaca pola capaian akademik siswa pada berbagai mata pelajaran dan membandingkannya dengan siswa lain yang memiliki karakteristik nilai serupa. Dataset tersebut kemudian digunakan dalam proses pengujian model dengan beberapa skenario pembagian data dan nilai k.

3.5. Rincian Perhitungan KNN Menggunakan Data Sampel

Pada bagian ini disajikan contoh perhitungan algoritma K-Nearest Neighbor (KNN) menggunakan data sampel nilai akademik siswa pada semester Ganjil tahun akademik 2025/2026. Contoh ini bertujuan untuk memperlihatkan alur kerja KNN dalam mengklasifikasikan siswa ke dalam kategori Berpotensi Lulus atau Berisiko Tidak Lulus berdasarkan kedekatan pola nilai mata pelajaran.

Pada implementasi sistem e-Rapor, atribut yang digunakan dalam proses KNN berasal dari seluruh nilai rapor mata pelajaran yang valid. Nilai yang kosong atau ditampilkan dengan tanda “-” tidak digunakan dalam proses perhitungan karena menandakan bahwa mata pelajaran tersebut tidak berlaku atau belum memiliki nilai pada siswa tertentu. Untuk menjaga keterbacaan perhitungan manual, contoh pada bagian ini hanya menggunakan enam atribut mata pelajaran yang memiliki data nilai valid, yaitu Matematika, Bahasa Indonesia, Bahasa Inggris, Pendidikan Pancasila, Kimia, dan Geografi. Sementara itu, pada sistem yang berjalan, proses klasifikasi tetap menggunakan seluruh mata pelajaran valid yang tersedia pada dataset.

Label acuan ditentukan berdasarkan aturan akademik yang digunakan pada sistem, yaitu siswa dikategorikan Berpotensi Lulus apabila rata-rata nilai mata pelajaran valid minimal 80 dan tidak terdapat nilai mata pelajaran valid di bawah 80. Sebaliknya, siswa dikategorikan Berisiko Tidak Lulus apabila rata-rata nilai mata pelajaran valid kurang dari 80 atau terdapat salah satu nilai mata pelajaran valid di bawah 80.

Tabel 1. Data Latih Sampel KNN

NIS	Nama Siswa	Matematika	B. Indonesia	B. Inggris	Pancasila	Kimia	Geografi	Rata-rata	Nilai Terendah	Label
0250**	Amelia Randa Bala Lembang	86	88	90	92	83	82	86,83	82	Berpotensi Lulus
0250**	Angelin Calista S.L	83	88	89	80	81	82	83,83	80	Berpotensi Lulus

NIS	Nama Siswa	Matematika	B. Indonesia	B. Inggris	Pancasila	Kimia	Geografi	Rata-rata	Nilai Terendah	Label
0250**	Aquene Natania Palamba	81	87	90	83	74	76	81,83	74	Berisiko Tidak Lulus
0250**	Bintang Dwi Patintingan	82	88	88	81	81	82	83,67	81	Berpotensi Lulus
0250**	Brodi Haydan Zanin	77	84	85	80	79	73	79,67	73	Berisiko Tidak Lulus
0250**	Darrent Casey Sumaryoto Malluka	81	85	87	84	80	81	83,00	80	Berpotensi Lulus
0250**	Farrel Yustra Nathanael	78	84	89	84	79	80	82,33	78	Berisiko Tidak Lulus
0250**	Forlan Afselio Rantetasak	80	85	89	83	79	80	82,67	79	Berisiko Tidak Lulus
0250**	Galuh Eka Kirana	83	88	90	88	85	84	86,33	83	Berpotensi Lulus
0250**	Inez Imelda Rottie	80	88	90	87	86	80	85,17	80	Berpotensi Lulus

Data uji yang digunakan dalam contoh perhitungan ini adalah salah satu siswa dengan kode NIS 0250**. Data uji ini dipilih untuk menunjukkan bagaimana algoritma KNN bekerja dalam menentukan status prediksi berdasarkan kemiripan pola nilai dengan data latih. Melalui proses tersebut, sistem menghitung jarak antara nilai mata pelajaran pada data uji dan data latih, kemudian menentukan kelas prediksi berdasarkan mayoritas tetangga terdekat. Dengan demikian, contoh ini memperlihatkan bahwa hasil prediksi tidak ditentukan hanya dari satu nilai, tetapi dari kedekatan keseluruhan pola nilai akademik siswa terhadap data pembandingan yang tersedia.

Tabel 2. Data Uji

NIS	Nama Siswa	Matematika	B. Indonesia	B. Inggris	Pancasila	Kimia	Geografi	Rata-rata	Nilai Terendah	Label
0250**	Cecilia Frilita Pabeo	92	90	89	91	92	84	89,67	84	Berpotensi Lulus

Sebelum dilakukan perhitungan jarak, setiap atribut numerik dinormalisasi menggunakan metode Min-Max. Normalisasi diperlukan karena KNN sensitif terhadap skala nilai. Dengan normalisasi, seluruh atribut berada pada skala yang sebanding sehingga tidak ada satu atribut yang memberikan pengaruh terlalu dominan hanya karena memiliki rentang nilai yang lebih besar.

Rumus normalisasi Min-Max adalah sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan:

X' = nilai hasil normalisasi

X = nilai asli

Xmin = nilai minimum pada atribut data latih

Xmax = nilai maksimum pada atribut data latih

Rentang nilai minimum dan maksimum dari data latih dapat dilihat pada Tabel 3.

Tabel 3. Nilai Minimum dan Maksimum Atribut Data Latih

Atribut	Nilai Minimum	Nilai Maksimum
Matematika	77	86
Bahasa Indonesia	84	88
Bahasa Inggris	85	90
Pendidikan Pancasila	80	92
Kimia	74	86

Atribut	Nilai Minimum	Nilai Maksimum
Geografi	73	84

Setelah data dinormalisasi, tahap berikutnya adalah menghitung jarak antara data uji dan setiap data latih menggunakan Euclidean Distance. Rumus Euclidean Distance adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$ = jarak antara data uji dan data latih

x_i = nilai atribut ke- i pada data uji

y_i = nilai atribut ke- i pada data latih

n = jumlah atribut yang digunakan

Hasil perhitungan jarak data uji terhadap data latih setelah normalisasi ditunjukkan pada Tabel 4.

Tabel 4. Pengurutan Jarak Terdekat

No	NIS	Nama Siswa	Label Data Latih	Jarak
1	0250**	Amelia Randa Bala Lembang	Berpotensi Lulus	1,1563
2	0250**	Galuh Eka Kirana	Berpotensi Lulus	1,3011
3	0250**	Inez Imelda Rottie	Berpotensi Lulus	1,6004
4	0250**	Angelin Calista S.L	Berpotensi Lulus	1,7215
5	0250**	Bintang Dwi Patintingan	Berpotensi Lulus	1,7585
6	0250**	Darrent Casey Sumaryoto Malluka	Berpotensi Lulus	2,1520
7	0250**	Forlan Afselio Rantetasak	Berisiko Tidak Lulus	2,2562
8	0250**	Aquene Natania Palamba	Berisiko Tidak Lulus	2,3064
9	0250**	Farrel Yustra Nathanael	Berisiko Tidak Lulus	2,5131
10	0250**	Brodi Haydan Zanin	Berisiko Tidak Lulus	2,9465

Berdasarkan hasil perhitungan jarak, data dengan nilai jarak terkecil dianggap memiliki tingkat kemiripan paling tinggi dengan data uji. Semakin kecil nilai jarak yang dihasilkan, semakin dekat pola nilai akademik data latih terhadap pola nilai akademik data uji. Sebaliknya, semakin besar nilai jarak, semakin rendah tingkat kemiripan antara kedua data tersebut. Setelah seluruh jarak antara data uji dan data latih dihitung, sistem mengurutkan hasil perhitungan dari jarak terkecil hingga terbesar untuk menentukan data latih yang paling relevan sebagai pembanding. Selanjutnya, sistem mengambil sejumlah tetangga terdekat sesuai dengan nilai k yang digunakan. Dalam penelitian ini, nilai k yang diuji adalah 3, 5, dan 7. Pengujian beberapa nilai k dilakukan untuk melihat pengaruh jumlah tetangga terhadap hasil klasifikasi, sehingga dapat diketahui apakah prediksi yang dihasilkan tetap konsisten ketika jumlah data pembanding terdekat berubah.

Tabel 5. Hasil Klasifikasi Berdasarkan Nilai k

Nilai k	Tetangga Terdekat	Berpotensi Lulus	Berisiko Tidak Lulus	Hasil Prediksi
3	3 data terdekat	3	0	Berpotensi Lulus
5	5 data terdekat	5	0	Berpotensi Lulus
7	7 data terdekat	6	1	Berpotensi Lulus

Hasil voting menunjukkan bahwa pada nilai $k = 3$, seluruh tetangga terdekat berada pada kategori Berpotensi Lulus. Pada nilai $k = 5$, hasil yang diperoleh juga sama, yaitu seluruh tetangga terdekat berada pada kategori Berpotensi Lulus. Sementara itu, pada nilai $k = 7$, terdapat enam tetangga dengan kategori Berpotensi Lulus dan satu tetangga dengan kategori Berisiko Tidak Lulus. Karena kelas mayoritas adalah Berpotensi Lulus, maka data uji dengan NIS 0250** diklasifikasikan sebagai Berpotensi Lulus.

Dengan demikian, berdasarkan contoh perhitungan manual menggunakan data sampel, algoritma KNN mampu menentukan status prediksi siswa berdasarkan kedekatan pola nilai mata pelajaran. Hasil ini sejalan dengan label acuan pada data uji, karena siswa tersebut memiliki rata-rata nilai 89,67 dan nilai terendah

84, sehingga memenuhi kriteria Berpotensi Lulus.

3.6. Evaluasi Model KNN dan Confusion Matrix

Evaluasi model dilakukan untuk mengetahui kemampuan algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan status akademik siswa ke dalam dua kategori, yaitu Berpotensi Lulus dan Berisiko Tidak Lulus. Evaluasi dilakukan menggunakan dataset nilai akhir siswa semester Ganjil tahun akademik 2025/2026 yang telah melalui proses seleksi, pembersihan, dan validasi nilai. Dataset yang digunakan berjumlah 283 data siswa.

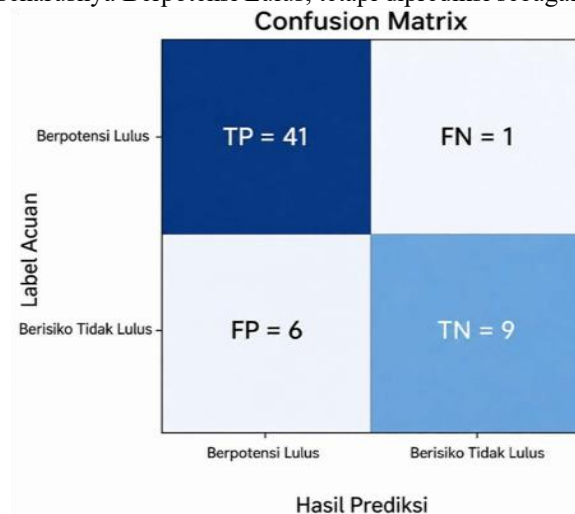
Pada tahap evaluasi, sistem menguji beberapa skenario pembagian data, yaitu 70:30, 75:25, 80:20, 85:15, dan 90:10, dengan nilai $k = 3$, $k = 5$, dan $k = 7$. Pemilihan hasil terbaik dilakukan berdasarkan urutan prioritas metrik, yaitu accuracy, F1-score, recall, precision, jumlah prediksi benar, jumlah prediksi salah paling sedikit, dan nilai k apabila terdapat hasil yang sama.

Berdasarkan hasil pengujian, skenario terbaik diperoleh pada pembagian data 80:20 dengan nilai $k = 3$. Dari 283 data, sebanyak 226 data digunakan sebagai data latih dan 57 data digunakan sebagai data uji. Model menghasilkan 50 prediksi benar dan 7 prediksi salah. Nilai accuracy yang diperoleh adalah 87,72%, precision sebesar 87,23%, recall sebesar 97,62%, dan F1-score sebesar 92,13%.

Tabel 6. Hasil Klasifikasi Berdasarkan Nilai k

Split Data	Nilai k	Jumlah Data	Data Latih	Data Uji	Prediksi Benar	Prediksi Salah	Accuracy	Precision	Recall	F1-Score
80:20	3	283	226	57	50	7	87,72%	87,23%	97,62%	92,13%

Confusion matrix digunakan untuk melihat rincian hasil klasifikasi secara lebih jelas. Pada hasil split terbaik, diperoleh nilai TP sebanyak 41, TN sebanyak 9, FP sebanyak 6, dan FN sebanyak 1. Nilai TP menunjukkan jumlah siswa yang benar diklasifikasikan sebagai Berpotensi Lulus, sedangkan TN menunjukkan jumlah siswa yang benar diklasifikasikan sebagai Berisiko Tidak Lulus. FP menunjukkan siswa yang seharusnya masuk kategori Berisiko Tidak Lulus, tetapi diprediksi sebagai Berpotensi Lulus. FN menunjukkan siswa yang seharusnya Berpotensi Lulus, tetapi diprediksi sebagai Berisiko Tidak Lulus.



Gambar 2. Confusion Matrix hasil pengujian KNN pada split data 80:20 dan nilai $k = 3$

Rumus accuracy yang digunakan adalah sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Accuracy = \frac{41 + 9}{41 + 9 + 6 + 1} \times 100\% = 87,72\%$$

Precision dihitung sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{41}{41 + 6} \times 100\% = 87,23\%$$

Recall dihitung sebagai berikut:

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

$$Recall = \frac{41}{41 + 1} \times 100\% = 97,62\%$$

F1-score dihitung sebagai berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1 - Score = 2 \times \frac{87,23\% \times 97,62\%}{87,23\% + 97,62\%} = 92,13\%$$

Hasil tersebut menunjukkan bahwa model KNN mampu mengklasifikasikan sebagian besar data uji dengan benar. Nilai recall sebesar 97,62% menunjukkan bahwa model sangat baik dalam mengenali siswa yang termasuk kategori Berpotensi Lulus. Sementara itu, adanya 6 data FP dan 1 data FN menunjukkan bahwa masih terdapat beberapa pola nilai siswa yang berdekatan dengan kelas berbeda, sehingga model belum sepenuhnya bebas dari kesalahan klasifikasi. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa KNN dapat digunakan sebagai alat bantu analisis akademik pada sistem e-Rapor, khususnya untuk mendukung identifikasi awal siswa yang berpotensi lulus maupun berisiko tidak lulus secara akademik.

3.7. Pengujian Black Box Sistem

Pengujian Black Box dilakukan untuk memastikan seluruh fitur utama sistem e-Rapor berjalan sesuai kebutuhan pengguna. Pengujian berfokus pada keluaran sistem berdasarkan input yang diberikan, tanpa meninjau struktur kode program. Fitur yang diuji meliputi autentikasi pengguna, pengelolaan data master, pengelolaan tahun akademik, input penilaian, pengolahan nilai akhir, prediksi akademik menggunakan KNN, serta pencetakan dokumen rapor.

Tabel 7. Hasil Pengujian Black Box

No	Fitur yang Diuji	Hasil yang Diharapkan	Hasil Pengujian	Keterangan
1	Login dan pilih peran	Sistem memvalidasi pengguna dan menampilkan halaman sesuai hak akses	Sesuai	Berhasil
2	Kelola data siswa	Sistem dapat menambah, mengubah, menghapus, dan menampilkan data siswa	Sesuai	Berhasil
3	Kelola tahun akademik	Sistem dapat mengaktifkan satu tahun akademik dan semester sebagai periode aktif	Sesuai	Berhasil
4	Kelola mata pelajaran	Sistem dapat menampilkan dan mengelola data mata pelajaran yang digunakan dalam penilaian	Sesuai	Berhasil
5	Input asesmen sumatif	Sistem dapat menyimpan nilai sumatif dan menghitung nilai rapor siswa	Sesuai	Berhasil
6	Rekap nilai akhir	Sistem dapat menampilkan rekap nilai akhir siswa berdasarkan kelas dan mata pelajaran	Sesuai	Berhasil
7	Simpan nilai akhir	Sistem dapat menyimpan nilai akhir dan nilai mata pelajaran valid ke database	Sesuai	Berhasil
8	Proses prediksi akademik KNN	Sistem dapat memproses data nilai menggunakan algoritma KNN	Sesuai	Berhasil
9	Evaluasi model KNN	Sistem dapat menampilkan accuracy, precision, recall, F1-score, dan confusion matrix	Sesuai	Berhasil
10	Cetak rapor	Sistem dapat menampilkan dan mencetak rapor sesuai data nilai siswa	Sesuai	Berhasil

Berdasarkan hasil pengujian Black Box, seluruh fitur utama sistem e-Rapor dapat berjalan sesuai dengan kebutuhan. Sistem mampu mengelola data akademik, memproses nilai, menyimpan rekap nilai akhir, menjalankan prediksi akademik menggunakan algoritma KNN, serta menampilkan dokumen rapor untuk kebutuhan cetak. Hasil ini menunjukkan bahwa sistem e-Rapor berbasis web telah berfungsi sebagai sistem

pengelolaan akademik sekaligus alat bantu analisis prediksi akademik siswa.

3.8. Implementasi Modul Prediksi Akademik

Modul Prediksi Akademik diimplementasikan sebagai bagian dari pengembangan sistem e-Rapor untuk memanfaatkan data nilai siswa sebagai dasar analisis prediktif. Modul ini tidak ditempatkan sebagai pengganti proses penilaian atau penentu akhir kelulusan, tetapi sebagai fitur pendukung yang membantu pihak sekolah mengenali siswa yang Berpotensi Lulus dan siswa yang Berisiko Tidak Lulus secara akademik. Dengan demikian, hasil prediksi digunakan sebagai informasi awal untuk mendukung pemantauan dan evaluasi akademik siswa.

Sumber data utama pada modul ini berasal dari fitur Nilai Akhir. Fitur Nilai Akhir menampilkan rekap nilai rapor siswa berdasarkan kelas dan mata pelajaran. Setiap nilai mata pelajaran yang valid disimpan ke dalam database dan digunakan sebagai atribut dalam proses klasifikasi KNN. Nilai yang kosong atau ditampilkan dengan tanda “-” tidak digunakan dalam perhitungan karena menunjukkan bahwa mata pelajaran tersebut tidak berlaku atau belum memiliki nilai pada siswa tertentu. Dengan pendekatan ini, proses prediksi tidak hanya bergantung pada satu nilai akhir, tetapi mempertimbangkan pola nilai siswa pada setiap mata pelajaran yang tersedia.

Proses prediksi diawali dengan pengambilan data nilai siswa berdasarkan tahun akademik dan semester yang dipilih, yaitu tahun akademik 2025/2026 semester Ganjil. Data tersebut kemudian melalui tahap preprocessing, meliputi seleksi data, validasi nilai numerik, penghapusan data duplikat, dan pencocokan data nilai dengan identitas siswa. Setelah data valid diperoleh, sistem membentuk dataset KNN berdasarkan nilai mata pelajaran yang memenuhi kriteria sebagai atribut numerik.

Label acuan pada sistem dibentuk menggunakan aturan akademik yang telah ditentukan. Siswa dikategorikan Berpotensi Lulus apabila rata-rata nilai mata pelajaran valid minimal 80 dan tidak terdapat nilai mata pelajaran valid di bawah 80. Sebaliknya, siswa dikategorikan Berisiko Tidak Lulus apabila rata-rata nilai mata pelajaran valid kurang dari 80 atau terdapat salah satu nilai mata pelajaran valid di bawah 80. Aturan ini digunakan sebagai dasar pembentukan kelas pada dataset, sedangkan proses klasifikasi dilakukan menggunakan algoritma KNN berdasarkan kedekatan pola nilai antar siswa.

Pada tahap klasifikasi, sistem melakukan normalisasi Min-Max terhadap seluruh atribut numerik agar setiap nilai mata pelajaran berada pada skala yang sebanding. Setelah itu, jarak antara data uji dan data latih dihitung menggunakan Euclidean Distance. Data latih dengan jarak terkecil dianggap sebagai data yang paling mirip dengan data uji. Selanjutnya, sistem menentukan hasil prediksi berdasarkan mayoritas label dari sejumlah tetangga terdekat sesuai nilai k yang digunakan. Pengujian model dilakukan menggunakan beberapa skenario pembagian data dan nilai k agar performa model dapat dianalisis secara lebih menyeluruh.

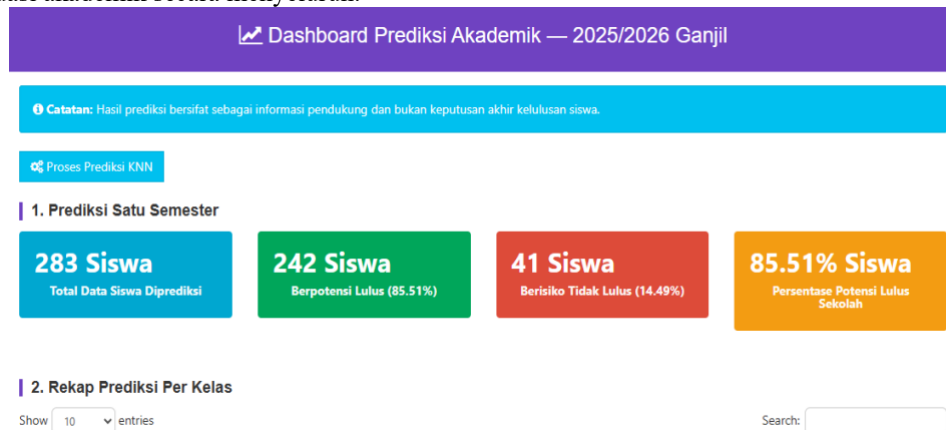
Hasil prediksi ditampilkan pada Dashboard Prediksi Akademik. Dashboard ini terdiri atas beberapa komponen utama. Komponen pertama adalah Prediksi Satu Semester, yang menampilkan jumlah siswa yang diproses, jumlah siswa Berpotensi Lulus, jumlah siswa Berisiko Tidak Lulus, serta persentase potensi lulus sekolah. Berdasarkan hasil implementasi pada data tahun akademik 2025/2026 semester Ganjil, sistem memproses 283 siswa, dengan 242 siswa Berpotensi Lulus dan 41 siswa Berisiko Tidak Lulus. Persentase siswa yang Berpotensi Lulus adalah 85,51%.

Komponen kedua adalah Rekap Prediksi Per Kelas, yang menyajikan hasil agregasi prediksi berdasarkan kelas. Bagian ini membantu pihak sekolah melihat persebaran siswa Berpotensi Lulus dan Berisiko Tidak Lulus pada setiap kelas. Komponen ketiga adalah Detail Prediksi Seluruh Siswa, yang menampilkan daftar siswa beserta status prediksi dan nilai confidence. Komponen keempat adalah Detail Prediksi Per Kelas, yang memudahkan wali kelas memantau kondisi akademik siswa dalam kelas yang diampu.

Komponen kelima adalah Statistik Prediksi Per Kelas, yang menampilkan komposisi atau persentase hasil prediksi pada masing-masing kelas. Komponen keenam adalah Evaluasi Model KNN dan Confusion Matrix, yang digunakan untuk menilai performa model melalui metrik accuracy, precision, recall, F1-score, TP, TN, FP, dan FN. Komponen ketujuh adalah Riwayat Prediksi Per Semester, yang menyimpan hasil prediksi berdasarkan periode akademik yang pernah diproses sehingga pihak sekolah dapat membandingkan kondisi akademik siswa antarsemester.

Implementasi modul Prediksi Akademik memberikan nilai tambah pada sistem e-Rapor karena data nilai yang sebelumnya hanya berfungsi sebagai laporan akademik dapat dimanfaatkan sebagai dasar analisis berbasis data. Modul ini dapat membantu wali kelas, guru, dan pihak sekolah dalam mengidentifikasi siswa yang membutuhkan perhatian akademik lebih awal. Namun, hasil prediksi tetap bersifat sebagai informasi pendukung karena keputusan kelulusan resmi tetap bergantung pada kebijakan sekolah, ketentuan kurikulum,

dan evaluasi akademik secara menyeluruh.



Gambar 3. Tampilan dashboard Prediksi Akademik tahun akademik 2025/2026 semester Ganjil

3.9. Pembahasan

Hasil pengembangan sistem menunjukkan bahwa e-Rapor berbasis web tidak hanya dapat digunakan sebagai sarana administrasi nilai, tetapi juga dapat dikembangkan menjadi media analisis akademik berbasis data. Sistem memiliki pembagian hak akses berdasarkan peran pengguna, sehingga setiap aktor hanya dapat menjalankan fungsi sesuai dengan tanggung jawabnya. Admin mengelola data dasar, guru mata pelajaran menginput nilai akademik, guru BK mengelola absensi dan sikap, wali kelas memantau nilai akhir serta mencetak laporan, sedangkan wali murid memperoleh akses untuk melihat informasi hasil belajar siswa. Struktur akses tersebut membuat proses pengelolaan data akademik menjadi lebih terarah dan terkontrol.

Data nilai yang diinputkan oleh guru mata pelajaran menjadi sumber utama dalam pembentukan nilai akhir siswa. Nilai akhir tersebut kemudian disimpan dan dimanfaatkan sebagai dataset pada Modul Prediksi Akademik. Dengan demikian, proses prediksi tidak berdiri sebagai fungsi yang terpisah dari sistem, tetapi terintegrasi dengan alur akademik yang sudah berjalan, mulai dari input nilai, rekap nilai akhir, penyimpanan nilai mata pelajaran valid, hingga proses klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN).

Penerapan KNN pada Modul Prediksi Akademik menunjukkan bahwa data nilai siswa dapat dimanfaatkan untuk menghasilkan informasi prediktif. Dalam sistem yang dikembangkan, klasifikasi dilakukan berdasarkan pola nilai mata pelajaran valid, bukan hanya berdasarkan satu nilai akhir. Siswa dikategorikan Berpotensi Lulus apabila rata-rata nilai mata pelajaran valid minimal 80 dan tidak terdapat nilai mata pelajaran valid di bawah 80. Sebaliknya, siswa dikategorikan Berisiko Tidak Lulus apabila rata-rata nilai mata pelajaran valid kurang dari 80 atau terdapat salah satu nilai mata pelajaran valid di bawah 80. Aturan tersebut digunakan sebagai label acuan, sedangkan KNN digunakan untuk mengenali kedekatan pola nilai antar siswa melalui proses normalisasi Min-Max dan perhitungan Euclidean Distance.

Berdasarkan hasil implementasi pada data tahun akademik 2025/2026 semester Ganjil, sistem memproses 283 data siswa. Dari jumlah tersebut, 242 siswa diklasifikasikan sebagai Berpotensi Lulus dan 41 siswa diklasifikasikan sebagai Berisiko Tidak Lulus, dengan persentase potensi lulus sebesar 85,51%. Hasil ini menunjukkan bahwa sebagian besar siswa memiliki capaian akademik yang memenuhi batas nilai yang digunakan dalam sistem, sementara sebagian lainnya perlu memperoleh perhatian lebih karena memiliki rata-rata nilai rendah atau masih terdapat nilai mata pelajaran valid di bawah 80.

Evaluasi model KNN dilakukan menggunakan beberapa skenario pembagian data dan nilai k. Hasil split data terbaik diperoleh pada skenario 80:20 dengan nilai $k = 3$. Dari 283 data, sebanyak 226 data digunakan sebagai data latih dan 57 data sebagai data uji. Model menghasilkan 50 prediksi benar dan 7 prediksi salah, dengan accuracy sebesar 87,72%, precision sebesar 87,23%, recall sebesar 97,62%, dan F1-score sebesar 92,13%. Nilai recall yang tinggi menunjukkan bahwa model sangat baik dalam mengenali siswa yang termasuk kategori Berpotensi Lulus. Namun, masih terdapat 6 data False Positive dan 1 data False Negative, yang menunjukkan bahwa terdapat beberapa siswa dengan pola nilai yang berdekatan dengan kelas berbeda sehingga model belum sepenuhnya bebas dari kesalahan klasifikasi.

Hasil confusion matrix memperlihatkan bahwa model menghasilkan 41 True Positive dan 9 True Negative. Artinya, model mampu mengklasifikasikan sebagian besar data uji secara tepat, baik pada kategori Berpotensi Lulus maupun Berisiko Tidak Lulus. Namun, nilai False Positive yang berjumlah 6 menunjukkan bahwa terdapat beberapa siswa yang seharusnya Berisiko Tidak Lulus, tetapi diprediksi Berpotensi Lulus.

Kondisi ini penting diperhatikan karena dalam konteks akademik, kesalahan membaca siswa berisiko dapat berdampak pada keterlambatan pemberian perhatian atau intervensi pembelajaran.

Secara praktis, Modul Prediksi Akademik dapat membantu wali kelas, guru, dan pihak sekolah dalam memantau kondisi akademik siswa secara lebih cepat. Informasi prediksi yang ditampilkan pada dashboard dapat digunakan sebagai dasar awal untuk mengidentifikasi siswa yang membutuhkan pendampingan, remedial, atau perhatian akademik tambahan. Dengan demikian, sistem ini tidak hanya berfungsi sebagai media pelaporan nilai, tetapi juga sebagai alat bantu peringatan dini akademik.

Meskipun demikian, hasil prediksi tetap harus dipahami sebagai informasi pendukung, bukan sebagai keputusan akhir kelulusan siswa. Keputusan kelulusan resmi tetap berada pada kebijakan sekolah dan harus mempertimbangkan ketentuan kurikulum, hasil evaluasi akademik, absensi, sikap, serta pertimbangan dewan guru. Oleh karena itu, kontribusi utama sistem ini adalah menyediakan dukungan analisis berbasis data agar pihak sekolah dapat melakukan pemantauan akademik secara lebih objektif, terukur, dan sistematis.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, sistem e-Rapor berbasis web berhasil dikembangkan untuk mendukung pengelolaan data akademik siswa sekaligus menyediakan Modul Prediksi Akademik menggunakan algoritma K-Nearest Neighbor (KNN). Modul ini memanfaatkan nilai rapor setiap mata pelajaran yang valid sebagai atribut klasifikasi, sehingga data akademik tidak hanya berfungsi sebagai laporan hasil belajar, tetapi juga dapat digunakan sebagai dasar analisis prediktif.

Pada implementasi sistem, data yang digunakan berasal dari tahun akademik 2025/2026 semester Ganjil dengan jumlah 283 data siswa. Hasil prediksi menunjukkan bahwa 242 siswa diklasifikasikan sebagai Berpotensi Lulus dan 41 siswa sebagai Berisiko Tidak Lulus, dengan persentase potensi lulus sebesar 85,51%. Status tersebut bersifat sebagai informasi pendukung untuk membantu pihak sekolah memantau kondisi akademik siswa secara lebih awal.

Evaluasi model KNN menunjukkan bahwa hasil split data terbaik diperoleh pada skenario 80:20 dengan nilai $k = 3$. Dari 283 data, sebanyak 226 data digunakan sebagai data latih dan 57 data sebagai data uji. Model menghasilkan 50 prediksi benar dan 7 prediksi salah, dengan accuracy sebesar 87,72%, precision sebesar 87,23%, recall sebesar 97,62%, dan F1-score sebesar 92,13%. Hasil confusion matrix menunjukkan 41 TP, 9 TN, 6 FP, dan 1 FN.

Pengujian Black Box menunjukkan bahwa fitur utama sistem, mulai dari login, pengelolaan data akademik, input nilai, rekap nilai akhir, proses prediksi akademik, evaluasi model, hingga cetak rapor dapat berjalan sesuai kebutuhan. Dengan demikian, sistem e-Rapor berbasis web yang dilengkapi KNN dapat digunakan sebagai alat bantu analisis akademik dan peringatan dini. Namun, hasil prediksi bukan merupakan keputusan akhir kelulusan siswa, karena keputusan resmi tetap harus mempertimbangkan kebijakan sekolah, ketentuan kurikulum, dan evaluasi akademik maupun non-akademik secara menyeluruh.

5. SARAN

Berdasarkan hasil penelitian, pengembangan berikutnya dapat dilakukan dengan memperluas dataset menggunakan data historis dari beberapa tahun akademik. Data yang lebih banyak dan beragam dapat membantu model mengenali pola akademik siswa secara lebih stabil dan representatif. Penelitian selanjutnya juga dapat menambahkan atribut non-akademik, seperti absensi, sikap, kedisiplinan, dan catatan wali kelas. Penambahan atribut tersebut penting karena potensi kelulusan siswa tidak hanya dipengaruhi oleh nilai akademik, tetapi juga oleh aspek kehadiran, perilaku, dan perkembangan siswa selama proses pembelajaran.

Selain itu, label klasifikasi sebaiknya dikembangkan menggunakan data kelulusan resmi sekolah apabila tersedia. Dengan menggunakan label resmi, proses evaluasi model dapat menjadi lebih kuat karena hasil prediksi dibandingkan dengan keputusan akademik yang benar-benar terjadi. Pengembangan sistem juga dapat dilakukan dengan membandingkan KNN dengan algoritma klasifikasi lain, seperti Naive Bayes, Decision Tree, Support Vector Machine, atau Random Forest. Perbandingan tersebut diperlukan untuk mengetahui metode yang paling sesuai dengan karakteristik data akademik pada sistem e-Rapor.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Dipa Makassar, khususnya Program Studi Teknik Informatika dan Rekayasa Perangkat Lunak yang telah memberikan dukungan dalam proses

penyusunan penelitian ini. Penulis juga mengucapkan terima kasih kepada dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan selama proses penyusunan paper ini.

Ucapan terima kasih juga disampaikan kepada SMA Kristen Elim Makassar yang telah menjadi objek studi kasus dalam penelitian ini serta memberikan dukungan terhadap pengembangan sistem e-Rapor berbasis web. Penulis juga berterima kasih kepada seluruh pihak yang telah membantu, baik secara langsung maupun tidak langsung, dalam proses pengumpulan data, pengembangan sistem, dan penyelesaian penelitian ini.

DAFTAR PUSTAKA

- [1] J. Jevri, S. Zakir, Z. Sesmiarni, and L. Efriyanti, "Perancangan E-Rapor Menggunakan Bahasa Pemrograman PHP/MySQL dengan Memanfaatkan E-Mail sebagai Pelaporan kepada Wali Murid," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 637–644, 2023, doi: 10.36040/jati.v7i1.6155.
- [2] F. R. Sari and A. Sabandi, "Efektivitas E-Rapor di SMK Negeri Kota Padang," *Journal of Practice Learning and Educational Development*, vol. 4, no. 3, pp. 207–213, 2024, doi: 10.58737/jpled.v4i3.364.
- [3] M. A. S. Wahyudi, Sumarno, and Y. Rahmawati, "Sistem Informasi Raport Digital SD Berbasis Web pada Sekolah Dasar Negeri Kemiri," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 1927–1935, 2024, doi: 10.29100/jupi.v9i4.5482.
- [4] P. Mahendra, R. Agustina, and Y. S. Dwanoko, "Rancang Bangun E-Rapor Berbasis Web di SMP Islam Annuriyah Malang," *Journal of Information Technology, Information Systems and Communications*, vol. 2, no. 1, pp. 41–52, 2024, doi: 10.21067/jistic.v2i1.9870.
- [5] E. Febriyanto, R. S. Naufal, and Sulistiawati, "Planning of the Web-Based E-Raport Assessment System," *APTISI Transactions on Technopreneurship (ATT)*, vol. 2, no. 1, pp. 48–58, 2020, doi: 10.34306/att.v2i1.27.
- [6] R. S. Daulay, "Analisis Kritis dan Pengembangan Algoritma K-Nearest Neighbor (KNN): Sebuah Tinjauan Literatur," *Jurnal Pendidikan Sains dan Komputer*, vol. 4, no. 2, pp. 131–141, 2024, doi: 10.47709/jpsk.v4i02.5055.
- [7] M. Affikri, M. C. Aruan, and M. Irsan, "Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Kelulusan Siswa di SD Negeri 1 Kedungsari," *Journal of Information System, Applied, Management, Accounting and Research (JISAMAR)*, vol. 9, no. 4, pp. 1388–1402, 2025, doi: 10.52362/jisamar.v9i4.2065.
- [8] M. Musthofa, D. Yunitasari, N. Nasikhin, and J. Wang, "K-Nearest Neighbors (K-NN) Algorithm Model in Predicting the Graduation Rate of Teacher Professional Education Students in Indonesia," *International Journal of Social Learning (IJS�)*, vol. 4, no. 3, pp. 291–310, 2024, doi: 10.47134/ijsl.v4i3.277.
- [9] A. Muhaimin, M. A. Hariyadi, and M. Imamudin, "Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 193–202, 2024, doi: 10.55338/jikoms.v7i1.2865.
- [10] S. Susanti, D. N. Manurung, L. J. C. Ginting, M. U. Nazha, and R. Siregar, "Kualifikasi Penentuan Kelulusan dan Analisis Penilaian Pendidikan Melalui Kemampuan Peserta Didik," *Semantik: Jurnal Riset Ilmu Pendidikan, Bahasa dan Budaya*, vol. 2, no. 3, pp. 43–50, 2024, doi: 10.61132/semantik.v2i3.756.