

Klasifikasi Minat Pendidikan Tinggi Siswa SMK Al-Ikhlas Losari Menggunakan *Naive Bayes* Berbasis *Streamlit*

Ravena Fate Kahna^{1*}, Bambang Irawan², Otong Saeful Bachri³

¹⁻³Teknik Informatika, Fakultas Teknik, Universitas Muhadi Setiabudi, Indonesia

*Korespondensi penulis: ravenafate@gmail.com

Abstract. *The transition of Vocational High School (SMK) students to higher education requires objective data-driven guidance services. This research aims to develop and implement a web-based classification system using the Naive Bayes algorithm to help map the higher education specialization of students of SMK Al-Ikhlas Losari. A quantitative approach is used with data obtained through questionnaires, interviews, observations, and school supporting data. The operational dataset consisted of 556 students, namely 203 Office Management (MP) students and 353 Motorcycle Engineering (TSM) students. The data was divided into 444 training data (80%) and 112 test data (20%), with categorical attributes transformed through one-hot encoding. The classification model is built using Multinomial Naive Bayes and is integrated into the Streamlit interface. The results of the evaluation showed 100.00% accuracy with a confusion matrix that showed that all test data was classified correctly. The system presents tables, graphs, classification reports, and recommendations in the field of higher education according to the major. MP students are directed to the fields of management, administration, business, accounting, and economics, while TSM students are directed to the fields of mechanical engineering, automotive, industry, and engineering. However, there is a label inconsistency between the implementation of the system and the conclusions of the study. Therefore, perfect accuracy needs to be interpreted as performance against operational targets, while subsequent research needs to explicitly define labels of interest to make the conclusions more valid.*

Keywords: *Data Mining; Higher Education Interest; Naive Bayes; Streamlit; Vocational Students.*

Abstrak. Transisi siswa Sekolah Menengah Kejuruan (SMK) menuju pendidikan tinggi memerlukan layanan bimbingan berbasis data yang objektif. Penelitian ini bertujuan mengembangkan dan mengimplementasikan sistem klasifikasi berbasis web menggunakan algoritma *Naive Bayes* untuk membantu pemetaan peminatan pendidikan tinggi siswa SMK Al-Ikhlas Losari. Pendekatan kuantitatif digunakan dengan data yang diperoleh melalui angket, wawancara, observasi, serta data pendukung sekolah. Dataset operasional terdiri atas 556 siswa, yakni 203 siswa Manajemen Perkantoran (MP) dan 353 siswa Teknik Sepeda Motor (TSM). Data dibagi menjadi 444 data latih (80%) dan 112 data uji (20%), dengan atribut kategorikal ditransformasikan melalui *one-hot encoding*. Model klasifikasi dibangun menggunakan *Multinomial Naive Bayes* dan diintegrasikan ke antarmuka *Streamlit*. Hasil evaluasi menunjukkan akurasi 100,00% dengan *confusion matrix* yang memperlihatkan seluruh data uji terklasifikasi benar. Sistem menyajikan tabel, grafik, *classification report*, serta rekomendasi bidang pendidikan tinggi sesuai jurusan. Siswa MP diarahkan pada bidang manajemen, administrasi, bisnis, akuntansi, dan ekonomi, sedangkan siswa TSM diarahkan pada bidang teknik mesin, otomotif, industri, dan rekayasa. Namun, terdapat ketidakkonsistenan label antara implementasi sistem dan kesimpulan penelitian. Oleh karena itu, akurasi sempurna perlu ditafsirkan sebagai kinerja terhadap target operasional, sementara penelitian berikutnya perlu menetapkan definisi label minat secara eksplisit agar simpulan lebih valid.

Kata kunci: Data Mining; Minat Pendidikan Tinggi; Naive Bayes; Siswa SMK; Streamlit.

1. LATAR BELAKANG

Pendidikan tinggi memiliki peran strategis dalam memperkuat kualitas sumber daya manusia, terutama ketika lulusan Sekolah Menengah Kejuruan (SMK) harus menentukan pilihan antara memasuki dunia kerja dan melanjutkan studi. Pendidikan lanjutan dapat memperluas kompetensi, memperdalam keahlian, serta meningkatkan peluang pengembangan karier. Namun, kondisi siswa SMK tidak seragam. Perbedaan capaian akademik, dukungan keluarga, kondisi ekonomi, motivasi, dan rencana karier dapat membentuk kecenderungan yang berbeda dalam menentukan pendidikan setelah lulus.

SMK Al-Ikhlas Losari menghadapi kebutuhan untuk memperoleh gambaran peminatan siswa secara lebih terukur. Naskah penelitian sumber menjelaskan bahwa pembinaan dan pendampingan siswa masih cukup banyak bertumpu pada pengamatan subjektif. Kondisi tersebut berpotensi membuat strategi bimbingan karier kurang spesifik terhadap karakteristik siswa. Pendekatan berbasis data diperlukan agar sekolah dapat mengenali pola siswa dan menyusun intervensi bimbingan yang lebih tepat.

Data mining menyediakan pendekatan untuk menemukan pola dari data dan mengubahnya menjadi informasi yang mendukung pengambilan keputusan. Dalam klasifikasi pendidikan, Naive Bayes banyak digunakan karena perhitungannya relatif sederhana, efisien, dan dapat bekerja pada data kategorikal. Algoritma ini menentukan kelas berdasarkan probabilitas posterior dengan menggabungkan probabilitas awal kelas dan peluang kemunculan atribut pada masing-masing kelas. Karakteristik tersebut menjadikan Naive Bayes relevan untuk sistem klasifikasi siswa yang memanfaatkan data akademik maupun data karakteristik individu.

Sejumlah penelitian terdahulu memperlihatkan pemanfaatan Naive Bayes dalam konteks pendidikan. Mafakhir dan Solichin (2020) menerapkan Naive Bayes Classifier untuk penjurusan siswa, sedangkan Wibowo et al. (2020) menggunakan metode yang sama untuk memprediksi kelanjutan studi siswa ke perguruan tinggi dan memperoleh akurasi 86,53%. Firdaus (2021) memanfaatkan Naive Bayes untuk penentuan jurusan mahasiswa teknologi informasi. Pada konteks pemilihan konsentrasi, Walangare dan Sujatmiko (2022) melaporkan sistem berbasis web yang menggunakan nilai akademik sebagai dasar klasifikasi. Perkasa dan Purwiantono (2023), Fikrillah dan Kurniadi (2023), serta Rasyid et al. (2023) juga menunjukkan bahwa Naive Bayes dapat digunakan untuk membangun sistem rekomendasi jurusan atau program studi.

Meskipun penelitian klasifikasi pendidikan telah berkembang, masih terdapat kebutuhan untuk menerapkan model pada konteks lokal sekolah dan mengintegrasikannya dengan antarmuka yang mudah digunakan oleh pengguna nonteknis. Penelitian ini menempatkan SMK Al-Ikhlas Losari sebagai objek penerapan serta mengintegrasikan model dengan Streamlit. Pengguna dapat mengunggah data Excel, memilih sheet, menentukan target dan fitur, menjalankan evaluasi, melihat confusion matrix dan classification report, serta memperoleh visualisasi hasil klasifikasi.

Kebaruan praktis penelitian terletak pada integrasi alur pengolahan data dan model klasifikasi ke dalam satu aplikasi web interaktif yang dapat mendukung kegiatan bimbingan siswa. Namun, artikel ini juga mempertahankan secara transparan kondisi data pada naskah

sumber. Naskah tersebut memuat perbedaan antara deskripsi label minat pada beberapa bagian dengan target jurusan yang terlihat pada implementasi akhir. Perbedaan ini penting dicatat agar hasil tidak ditafsirkan melampaui target yang benar-benar dievaluasi pada aplikasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan sistem klasifikasi berbasis Naive Bayes, menganalisis hasil evaluasi model pada data siswa SMK Al-Ikhlas Losari, dan mengimplementasikan model melalui Streamlit sebagai alat bantu penyajian rekomendasi pendidikan tinggi. Hasil penelitian diharapkan dapat menjadi informasi awal bagi sekolah, guru Bimbingan dan Konseling (BK), serta siswa dalam menyusun arah pendidikan lanjutan secara lebih objektif dan berbasis data.

2. KAJIAN TEORITIS

Data Mining dan Klasifikasi

Data mining merupakan proses mengekstraksi pola, hubungan, kecenderungan, dan pengetahuan dari kumpulan data. Tahapan pengolahan pada naskah sumber mencakup data cleansing, data integration, data selection, data transformation, data mining, pattern evaluation, dan knowledge representation. Dalam penelitian pendidikan, rangkaian tersebut diperlukan agar data siswa yang berasal dari berbagai sumber dapat diubah menjadi data yang konsisten dan layak dianalisis.

Klasifikasi termasuk supervised learning karena model dibangun dari data yang telah memiliki kelas atau target. Secara umum proses klasifikasi terdiri atas persiapan data, pembentukan model menggunakan data latih, pengujian menggunakan data uji, dan evaluasi kinerja. Pemisahan data latih dan data uji penting untuk menilai kemampuan model pada data yang tidak digunakan selama pembentukan model. Educational data mining telah banyak memanfaatkan pendekatan klasifikasi untuk memprediksi performa, risiko studi, retensi, maupun pilihan akademik siswa (Albreiki et al., 2021; Feng et al., 2022; Shafiq et al., 2022).

Naive Bayes dan Evaluasi Kinerja

Naive Bayes adalah algoritma klasifikasi probabilistik yang berlandaskan Teorema Bayes dan menggunakan asumsi independensi kondisional antarfitur. Untuk suatu kelas C dan atribut X , model menghitung nilai yang sebanding dengan $P(C) \times P(X|C)$, kemudian memilih kelas dengan probabilitas terbesar. Asumsi independensi membuat proses komputasi relatif ringan, sehingga algoritma sesuai untuk aplikasi yang membutuhkan pelatihan dan prediksi cepat. Pada data kategorikal, nilai kategori dapat ditransformasikan ke representasi numerik sebelum diproses oleh model.

Evaluasi klasifikasi dapat dilakukan menggunakan confusion matrix yang membandingkan kelas aktual dengan kelas hasil prediksi. Dari matriks tersebut dapat dihitung accuracy, precision, recall, dan F1-score. Accuracy menunjukkan proporsi prediksi yang benar terhadap seluruh data uji, sedangkan precision dan recall memberikan informasi yang lebih spesifik terhadap kualitas prediksi per kelas. F1-score menggabungkan precision dan recall dalam satu ukuran harmonik. Penggunaan beberapa metrik diperlukan agar evaluasi tidak hanya bergantung pada satu angka, khususnya ketika distribusi kelas tidak seimbang (Tharwat, 2021).

Minat Pendidikan Tinggi dan Rekomendasi Jurusan

Minat pendidikan tinggi menggambarkan kecenderungan individu untuk melanjutkan studi yang dapat terbentuk dari ketertarikan, perhatian, motivasi, informasi pendidikan, pengalaman, kemampuan akademik, dan rencana karier. Dalam layanan bimbingan, klasifikasi minat seharusnya tidak diposisikan sebagai keputusan final, melainkan sebagai informasi awal yang perlu dikombinasikan dengan pertimbangan guru, siswa, dan keluarga.

Rekomendasi jurusan atau rumpun ilmu merupakan tahap lanjutan yang memetakan karakteristik siswa ke alternatif bidang studi. Sistem berbasis data dapat membantu menyajikan rekomendasi secara konsisten, tetapi kualitas rekomendasi tetap bergantung pada validitas label, relevansi atribut, representativitas data, serta kesesuaian antara target model dan konsep yang ingin diukur. Karena itu, definisi kelas pada dataset menjadi komponen metodologis yang sangat penting.

Streamlit sebagai Media Implementasi

Streamlit adalah framework Python untuk membangun aplikasi data dan machine learning berbasis browser. Naskah sumber memanfaatkan Streamlit untuk menghubungkan pengguna dengan proses pemuatan dataset, pemilihan target dan fitur, preprocessing, pelatihan model, evaluasi, prediksi, serta visualisasi. Pendekatan ini membuat hasil analisis lebih mudah diakses karena pengguna tidak harus berinteraksi langsung dengan kode program.

Pada penelitian ini, aplikasi menggunakan ekosistem Python dan pustaka pengolahan data serta machine learning. Dataset Excel dibaca ke dalam aplikasi, fitur kategorikal dikodekan menggunakan OneHotEncoder, model dilatih menggunakan Multinomial Naive Bayes, lalu performa ditampilkan dalam bentuk accuracy, confusion matrix, classification report, pie chart, bar chart, dan heatmap. Integrasi tersebut menempatkan aplikasi tidak hanya sebagai media demonstrasi, tetapi juga sebagai prototipe alat bantu pengambilan keputusan di lingkungan sekolah.

3. METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif karena data siswa diolah menjadi variabel kategorikal dan dianalisis menggunakan algoritma klasifikasi. Objek penelitian adalah siswa kelas akhir SMK Al-Ikhlas Losari. Pengumpulan data pada naskah sumber dilakukan melalui wawancara, angket berbasis Google Form, observasi, dan pemanfaatan data pendukung sekolah. Angket memuat informasi terkait nilai pelajaran, bakat, minat, dan hobi, sedangkan data sekolah digunakan untuk melengkapi karakteristik responden.

Dataset operasional yang digunakan dalam implementasi akhir berjumlah 556 siswa. Distribusi data berdasarkan tabel sampel pada naskah sumber terdiri atas tiga rombongan belajar Manajemen Perkantoran (MP) dan empat rombongan belajar Teknik Sepeda Motor (TSM), dengan 427 siswa laki-laki dan 129 siswa perempuan. Rincian dataset disajikan pada Tabel 1.

Tabel 1. Distribusi Data Siswa Kelas XII.

No.	Jurusan	Rombel	Laki-laki	Perempuan	Jumlah
1	MP / Manajemen Perkantoran	MP 1	35	32	67
2	MP / Manajemen Perkantoran	MP 2	30	30	60
3	MP / Manajemen Perkantoran	MP 3	38	38	76
4	TSM / Teknik Sepeda Motor	TSM 1	83	9	92
5	TSM / Teknik Sepeda Motor	TSM 2	80	10	90
6	TSM / Teknik Sepeda Motor	TSM 3	78	8	86
7	TSM / Teknik Sepeda Motor	TSM 4	83	2	85
Total			427	129	556

Dataset dibagi menggunakan train-test split dengan komposisi 80% data latih dan 20% data uji. Berdasarkan tampilan evaluasi pada aplikasi, pembagian tersebut menghasilkan 444 data latih dan 112 data uji. Pada implementasi akhir yang terdokumentasi, pengguna memilih kolom target “Jurusan”, sedangkan fitur kategorikal yang terlihat digunakan adalah “Rombel” dan “Jenis Kelamin”. Data kategorikal diproses menggunakan OneHotEncoder dengan penanganan kategori yang tidak dikenali, kemudian model Multinomial Naive Bayes dilatih pada data latih.

Alur analisis terdiri atas pemuatan dataset, pemilihan target dan fitur, preprocessing, pembagian data latih dan data uji, pelatihan model, prediksi, serta evaluasi. Sistem menghitung accuracy, confusion matrix, precision, recall, dan F1-score melalui classification report. Hasil klasifikasi selanjutnya dihubungkan dengan rekomendasi bidang pendidikan tinggi. Siswa MP diarahkan ke bidang manajemen, administrasi, bisnis, akuntansi, atau ekonomi, sedangkan siswa TSM diarahkan ke teknik mesin, teknik otomotif, teknik industri, atau bidang rekayasa.

Catatan konsistensi data diperlukan karena naskah sumber menyajikan beberapa deskripsi yang tidak sepenuhnya seragam. Pada bagian metode terdapat narasi “116 siswa MP dan 169 siswa TSM”, sedangkan Tabel 3.7, tampilan aplikasi, dan hasil akhir secara konsisten menunjukkan total 556 siswa dengan komposisi 203 MP dan 353 TSM. Artikel ini menggunakan angka 556 sebagai dataset operasional karena angka tersebut sesuai dengan tabel sampel dan implementasi sistem. Perbedaan tersebut tidak dihilangkan, melainkan diperlakukan sebagai keterbatasan dokumentasi yang perlu diperbaiki pada penelitian lanjutan.

4. HASIL DAN PEMBAHASAN

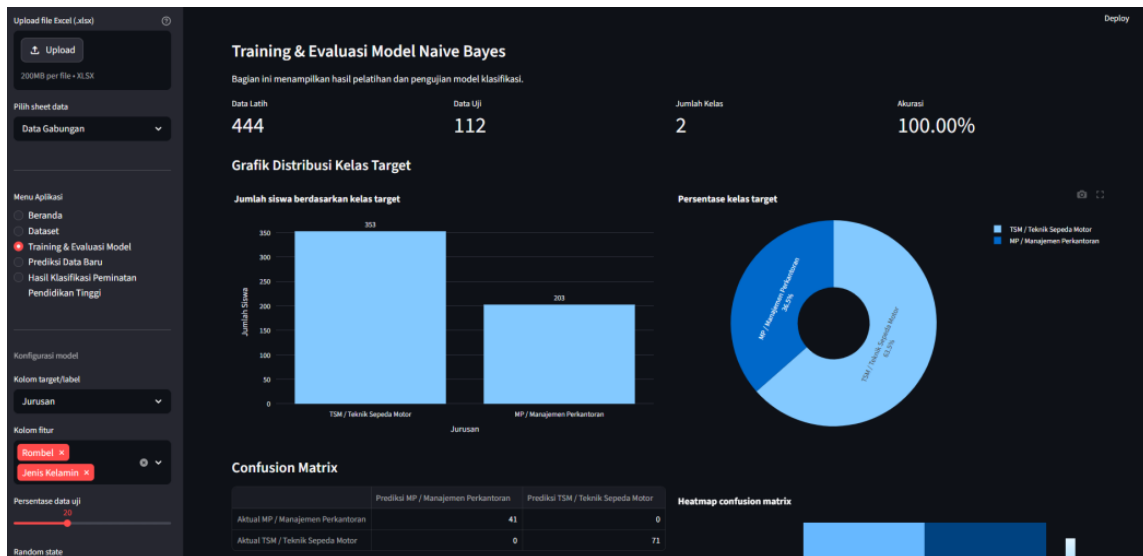
Implementasi Aplikasi dan Pengolahan Data

Aplikasi yang dikembangkan menempatkan seluruh tahapan klasifikasi dalam antarmuka Streamlit. Pengguna dapat mengunggah berkas Excel, memilih sheet, menentukan target, memilih fitur yang akan digunakan, dan menjalankan proses evaluasi. Jika berkas tidak diunggah, sistem dapat menggunakan dataset bawaan. Setelah data dibaca, fitur kategorikal dikonversi melalui one-hot encoding agar dapat diproses oleh model Multinomial Naive Bayes.

Desain ini menyederhanakan proses eksperimen karena pengguna dapat mengelola dataset dan melihat hasil tanpa menjalankan kode secara manual. Selain nilai evaluasi, sistem menyediakan grafik ringkasan dataset, distribusi kelas, confusion matrix, classification report, grafik prediksi, dan fasilitas untuk mengunduh hasil klasifikasi. Dengan demikian, keluaran model dapat dibaca secara lebih komunikatif oleh pihak sekolah.

Hasil Evaluasi Model

Tampilan evaluasi pada naskah sumber menunjukkan 444 data latih, 112 data uji, dua kelas, dan akurasi 100,00%. Target yang terlihat pada antarmuka adalah “Jurusan”. Confusion matrix pada data uji menunjukkan 41 data MP diprediksi sebagai MP dan 71 data TSM diprediksi sebagai TSM, sedangkan dua sel kesalahan bernilai nol. Dengan demikian, seluruh 112 data uji pada konfigurasi tersebut diklasifikasikan secara benar.



Gambar 1. Hasil Evaluasi Model Naive Bayes pada Aplikasi Streamlit.

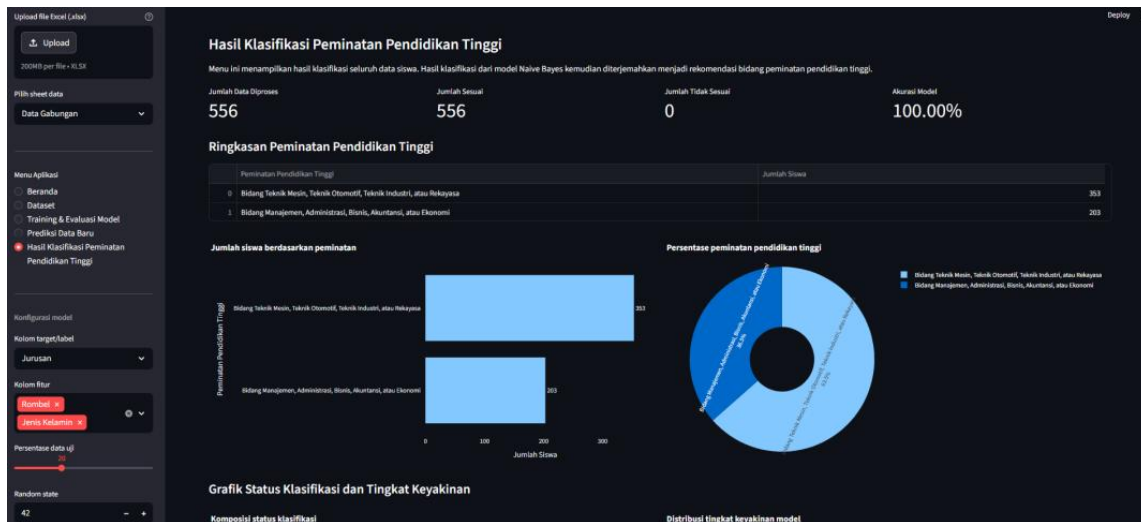
Akurasi sempurna menunjukkan bahwa fitur yang digunakan pada konfigurasi tersebut sangat kuat dalam membedakan kelas target. Namun, kondisi ini perlu dibaca secara hati-hati. Rombel pada dataset memiliki kode yang langsung berkaitan dengan jurusan, yaitu MP 1-MP 3 untuk kelas Manajemen Perkantoran dan TSM 1-TSM 4 untuk kelas Teknik Sepeda Motor. Karena target model adalah jurusan, penggunaan rombel sebagai fitur berpotensi membuat kelas target dapat dikenali hampir secara deterministik dari nama rombel. Secara metodologis, hal tersebut dapat menjelaskan akurasi 100% tanpa harus berarti bahwa model telah mengukur “minat melanjutkan pendidikan tinggi” secara sempurna. Temuan ini merupakan inferensi dari struktur fitur dan target yang terlihat pada implementasi, bukan perubahan terhadap hasil asli penelitian.

Temuan tersebut menegaskan pentingnya pemisahan antara variabel identitas kelas dan variabel substantif yang benar-benar merepresentasikan minat. Apabila tujuan utama penelitian adalah memprediksi minat melanjutkan studi, target sebaiknya berupa label minat yang dibentuk dari instrumen terukur, sedangkan fitur sebaiknya mencakup nilai akademik, dukungan sosial, rencana karier, kondisi ekonomi, motivasi, atau atribut lain yang tidak secara langsung mengkodekan target. Pendekatan ini akan menghasilkan pengujian generalisasi yang lebih bermakna.

Distribusi Klasifikasi dan Rekomendasi Bidang Studi

Halaman hasil menunjukkan bahwa 556 data berhasil diproses dan seluruhnya dinyatakan sesuai dengan klasifikasi model. Distribusi yang ditampilkan terdiri atas 353 siswa pada kelompok TSM dan 203 siswa pada kelompok MP. Sistem kemudian memetakan siswa TSM ke rekomendasi bidang teknik mesin, teknik otomotif, teknik industri, dan rekayasa,

sedangkan siswa MP diarahkan ke bidang manajemen, administrasi, bisnis, akuntansi, dan ekonomi.



Gambar 2. Hasil Klasifikasi dan Rekomendasi Peminatan Pendidikan Tinggi.

Distribusi tersebut sejalan dengan struktur sampel pada Tabel 1, yaitu jumlah siswa TSM lebih besar daripada MP. Rekomendasi yang dihasilkan bersifat pemetaan awal berdasarkan kelas/jurusan siswa dan tidak semestinya menggantikan proses konseling individual. Bagi sekolah, hasil dapat digunakan untuk menyiapkan informasi program studi yang lebih relevan dengan kompetensi keahlian siswa, menyusun kegiatan sosialisasi perguruan tinggi, serta menentukan kelompok siswa yang membutuhkan eksplorasi karier lebih lanjut.

Pembahasan terhadap Penelitian Terdahulu

Secara umum, hasil implementasi mendukung temuan penelitian terdahulu bahwa Naive Bayes dapat diterapkan pada data pendidikan. Wibowo et al. (2020) melaporkan akurasi 86,53% untuk prediksi kelanjutan studi, sedangkan Fikrillah dan Kurniadi (2023) memperoleh akurasi 73,4% pada sistem rekomendasi program studi. Walangare dan Sujatmiko (2022) melaporkan akurasi 100% pada pemilihan peminatan konsentrasi berbasis nilai akademik. Perbedaan akurasi antarpelitian menunjukkan bahwa kinerja model sangat bergantung pada struktur target, jumlah data, komposisi kelas, dan kekuatan hubungan antara fitur dengan kelas.

Penelitian lain juga menegaskan bahwa kualitas prediksi pendidikan dipengaruhi oleh pemilihan atribut dan desain evaluasi. Studi educational data mining memanfaatkan atribut akademik, aktivitas belajar, maupun karakteristik siswa untuk memprediksi performa dan risiko studi (Albreiki et al., 2021; Feng et al., 2022; Yağcı, 2022). Dengan demikian, pengembangan sistem di SMK Al-Ikhlas Losari akan lebih kuat apabila target “minat pendidikan tinggi” dibentuk secara eksplisit dan divalidasi, lalu dibandingkan dengan

algoritma lain seperti Decision Tree, K-Nearest Neighbor, dan Support Vector Machine sebagaimana juga disarankan pada naskah sumber.

Kekuatan utama penelitian ini adalah tersedianya prototipe aplikasi yang mengintegrasikan proses analisis dan visualisasi. Aplikasi mempermudah pengguna melihat data, menjalankan prediksi, dan mengunduh hasil. Keterbatasan utamanya adalah ketidakkonsistenan nomenklatur label antara beberapa bagian naskah sumber. Bab V menyebut 203 siswa sebagai “Minat” dan 353 siswa sebagai “Tidak minat”, sedangkan tampilan aplikasi dan distribusi data menunjukkan angka yang sama sebagai kelas MP dan TSM. Karena kedua interpretasi berasal dari sumber yang sama tetapi tidak ekuivalen secara konsep, artikel ini tidak menggabungkannya menjadi satu klaim. Angka 203 dan 353 dipertahankan sebagai distribusi kelas operasional pada aplikasi, sedangkan interpretasi Minat/Tidak minat dicatat sebagai bagian yang memerlukan validasi ulang terhadap dataset asli.

Implikasi Penelitian

Dari sisi terapan, sistem dapat menjadi alat bantu bagi guru BK dan bagian kesiswaan untuk menyajikan informasi peminatan secara cepat. Antarmuka Streamlit memungkinkan data baru dimasukkan tanpa harus mengubah kode, sehingga sistem dapat diperbarui ketika sekolah memperoleh data angkatan berikutnya. Visualisasi juga membantu pihak sekolah memahami distribusi data secara lebih intuitif.

Dari sisi metodologis, penelitian menunjukkan bahwa akurasi tinggi perlu selalu dikaji bersama struktur data. Penggunaan fitur yang mengandung informasi langsung tentang target dapat menghasilkan evaluasi yang sangat tinggi tetapi kurang mencerminkan kemampuan prediktif terhadap konsep yang ingin diukur. Oleh karena itu, validasi definisi target, pemeriksaan potensi data leakage, penggunaan cross-validation, dan pemisahan data berdasarkan skenario yang realistis perlu menjadi bagian pengembangan model berikutnya.

5. KESIMPULAN DAN SARAN

Penelitian menghasilkan prototipe sistem klasifikasi berbasis web menggunakan Multinomial Naive Bayes dan Streamlit untuk mengolah data siswa SMK Al-Ikhlas Losari. Dataset operasional pada implementasi akhir berjumlah 556 siswa dan dibagi menjadi 444 data latih serta 112 data uji. Pada konfigurasi aplikasi yang terdokumentasi, target model adalah jurusan dan fitur kategorikal yang digunakan mencakup rombel serta jenis kelamin. Evaluasi menampilkan akurasi 100,00% dengan 41 data MP dan 71 data TSM pada data uji seluruhnya diklasifikasikan dengan benar. Hasil akhir memetakan 203 siswa MP ke bidang manajemen, administrasi, bisnis, akuntansi, dan ekonomi, serta 353 siswa TSM ke bidang teknik mesin,

teknik otomotif, teknik industri, dan rekayasa. Sistem dapat membantu sekolah menyajikan gambaran awal arah pendidikan lanjutan secara interaktif, tetapi hasil tidak boleh diposisikan sebagai keputusan tunggal bagi siswa. Naskah sumber masih memuat ketidakkonsistenan antara label jurusan pada aplikasi dan label Minat/Tidak minat pada bagian kesimpulan, sehingga interpretasi akurasi terhadap “minat melanjutkan pendidikan tinggi” harus dilakukan secara hati-hati. Penelitian selanjutnya disarankan menetapkan target minat secara eksplisit melalui instrumen yang tervalidasi, menggunakan atribut yang lebih substantif, memeriksa potensi data leakage, memperbesar dan menyeimbangkan dataset, menambahkan cross-validation, serta membandingkan Naive Bayes dengan Decision Tree, K-Nearest Neighbor, Support Vector Machine, atau metode klasifikasi lain. Sistem juga dapat dikembangkan dengan fitur laporan dan pencetakan agar lebih mudah dimanfaatkan oleh guru BK dan pihak sekolah.

DAFTAR REFERENSI

- Ahmed, E. (2024). Student performance prediction using machine learning algorithms. *Applied Computational Intelligence and Soft Computing*, 2024, 4067721. <https://doi.org/10.1155/2024/4067721>
- Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of students' performance prediction using machine learning techniques. *Education Sciences*, 11(9), 552.
- Feng, G., Fan, M., & Chen, Y. (2022). Analysis and prediction of students' academic performance based on educational data mining. *IEEE Access*, 10, 19558–19571. <https://doi.org/10.1109/ACCESS.2022.3151652>
- Fikrillah, H. N. F., & Kurniadi, D. (2023). Rekomendasi pemilihan program studi menggunakan algoritma Naïve Bayes. *Jurnal Algoritma*, 20(1), 42–49. <https://doi.org/10.33364/algoritma/v.20-1.1236>
- Firdaus, K. L. (2021). Penerapan metode Naïve Bayes dalam prediksi penentuan jurusan mahasiswa TI. *Journal of Telecommunication Electronics and Control Engineering (JTECE)*, 2(2), 78–84. <https://doi.org/10.20895/jtece.v2i2.140>
- Hussain, S., & Khan, M. Q. (2023). Student-Performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of Data Science*, 10(3), 637–655. <https://doi.org/10.1007/s40745-021-00341-0>
- Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*, 7, 100204.
- Istighfar, F., Negara, A. B. P., & Tursina. (2023). Klasifikasi bidang keahlian mahasiswa menggunakan algoritma Naive Bayes. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 11(1). <https://doi.org/10.26418/justin.v11i1.52402>

- Kurniadi, D., Nuraeni, F., & Lestari, S. M. (2022). Implementasi algoritma Naïve Bayes menggunakan feature forward selection dan SMOTE untuk memprediksi ketepatan masa studi mahasiswa sarjana. *Jurnal Sistem Cerdas*, 5(2), 63–82.
- Mafakhir, A. Z., & Solichin, A. (2020). Penerapan metode Naïve Bayes Classifier untuk penjurusan siswa pada Madrasah Aliyah Al-Falah Jakarta. *Fountain of Informatics Journal*, 5(1). <https://doi.org/10.21111/fij.v5i1.4007>
- Mulyani, A., Kurniadi, D., Nashrulloh, M. R., Julianto, I. T., & Regita, M. (2022). The prediction of PPA and KIP-Kuliah scholarship recipients using Naive Bayes algorithm. *Jurnal Teknik Informatika (JUTIF)*, 3(4), 821–827. <https://doi.org/10.20884/1.jutif.2022.3.4.297>
- Munawirah, M., & Arisha, A. O. (2025). Implementasi Naïve Bayes untuk klasifikasi peminatan program studi pada penerimaan mahasiswa baru di Fakultas Ilmu Komputer Unika. *Bulletin of Information Technology (BIT)*, 6(3), 218–229. <https://doi.org/10.47065/bit.v6i3.2142>
- Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, 20(1), 4.
- Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials Today: Proceedings*, 80, 3782–3785. <https://doi.org/10.1016/j.matpr.2021.07.382>
- Perkasa, K. B. P. Y., & Purwiantono, F. E. (2023). Sistem rekomendasi jurusan menggunakan algoritma Naïve Bayes Gaussian berbasis web. *J-INTECH*, 11(2), 361–370. <https://doi.org/10.32664/j-intech.v11i2.1090>
- Rasyid, R. M. A. K., Riyanto, A., Widyawati, R., & Istiningsih. (2023). Implementasi algoritma Naïve Bayes untuk sistem rekomendasi pemilihan fakultas di Universitas Amikom Yogyakarta. *JIKOM: Jurnal Informatika dan Komputer*, 13(1), 1–9. <https://doi.org/10.55794/jikom.v13i1.93>
- Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student retention using educational data mining and predictive analytics: A systematic literature review. *IEEE Access*, 10, 72480–72503. <https://doi.org/10.1109/ACCESS.2022.3188767>
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Walangare, R. A. C., & Sujatmiko, B. (2022). Penerapan algoritma Naive Bayes dalam sistem pendukung keputusan pemilihan peminatan konsentrasi berdasarkan nilai akademik berbasis web pada Program Studi S1 Pendidikan Teknologi Informasi. *IT-Edu: Jurnal Information Technology and Education*, 7(3), 74–83. <https://doi.org/10.26740/it-edu.v7i3.50086>
- Wibowo, G. W. N., Arifin, Z., Romli, M. A., & Amal, N. I. (2020). Prediksi kelanjutan studi siswa ke perguruan tinggi dengan Naive Bayes. *Jurnal DISPROTEK*, 11(1), 41–46. <https://doi.org/10.34001/jdpt.v11i1.1159>
- Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11.