

Pemodelan Prediktif Lama Studi Mahasiswa Berbasis Kinerja Akademik Menggunakan Algoritma Naive Bayes

Predictive Modeling of Student Study Length Based on Academic Performance Using the Naive Bayes Algorithm

Irmawati¹; Firman Aziz^{2,*}; Ayu Lestari Aziz³

^{1,3} Irmex Digital Akademika, Makassar 90551, Indonesia

² Universitas Pancasakti, Makassar 90121, Indonesia

¹irmawati@irmexdigika.com; ²firman.aziz@unpacti.ac.id; ³ayulestariazis@irmexdigika.com

* Corresponding author

Abstrak

Ketepatan waktu penyelesaian studi mahasiswa merupakan indikator penting dalam evaluasi mutu pendidikan tinggi dan pengelolaan akademik. Keterlambatan kelulusan tidak hanya berdampak pada mahasiswa, tetapi juga memengaruhi kinerja institusi pendidikan. Penelitian ini bertujuan untuk mengembangkan model prediktif lama studi mahasiswa berbasis kinerja akademik menggunakan algoritma Naive Bayes. Data yang digunakan berupa data akademik mahasiswa yang mencakup Indeks Prestasi Semester (IPS) semester awal hingga menengah serta atribut pendukung lainnya. Tahapan penelitian meliputi pra-pemrosesan data, diskretisasi IPS, pembagian data latih dan data uji, pembangunan model klasifikasi, serta evaluasi kinerja model menggunakan confusion matrix dan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Naive Bayes mampu memprediksi lama studi mahasiswa dengan tingkat performa yang baik. Temuan penting menunjukkan bahwa stabilitas kinerja akademik memiliki pengaruh yang lebih signifikan dibandingkan capaian nilai akademik semata. Model yang dikembangkan berpotensi diterapkan sebagai sistem peringatan dini akademik untuk mengidentifikasi mahasiswa berisiko sejak tahap awal studi dan mendukung pengambilan kebijakan akademik berbasis data.

Kata Kunci: lama studi mahasiswa; kinerja akademik; Naive Bayes; data mining pendidikan; sistem peringatan dini

Abstract

Timely completion of study is a critical indicator of higher education quality and academic management effectiveness. Delayed graduation not only affects students but also impacts institutional performance. This study aims to develop a predictive model of student study duration based on academic performance using the Naive Bayes algorithm. The dataset consists of students' academic records, including semester-based Grade Point Averages (GPA) from early to mid-study periods and supporting attributes. The research methodology includes data preprocessing, GPA discretization, training-testing data splitting, model construction, and performance evaluation using a confusion matrix and classification metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that the Naive Bayes model achieves satisfactory predictive performance in classifying student study duration. An important finding indicates that the stability of academic performance plays a more significant role than absolute GPA values in predicting graduation timeliness. The proposed model has the potential to be implemented as an academic early warning system to identify at-risk students at an early stage and to support data-driven academic decision-making in higher education institutions.

Keywords: student study duration; academic performance; Naive Bayes; educational data mining; early warning system

Pendahuluan

Ketepatan waktu penyelesaian studi mahasiswa merupakan indikator penting dalam evaluasi mutu pendidikan tinggi dan kinerja institusi akademik. Lama studi mahasiswa berpengaruh langsung terhadap akreditasi program studi, efisiensi pembelajaran, serta reputasi perguruan tinggi [1]. Mahasiswa yang mengalami keterlambatan kelulusan berpotensi menghadapi hambatan akademik dan ekonomi, sementara institusi pendidikan dapat mengalami penurunan capaian indikator kinerja utama.

Keterlambatan penyelesaian studi mahasiswa umumnya dipengaruhi oleh berbagai faktor, salah satunya adalah kinerja akademik yang tercermin melalui Indeks Prestasi Semester (IPS). Beberapa penelitian menunjukkan bahwa IPS pada semester awal memiliki korelasi yang signifikan terhadap keberhasilan akademik mahasiswa pada tahap akhir studi [2]. Pola IPS yang rendah atau fluktuatif dapat menjadi indikator awal risiko ketidaktepatan waktu kelulusan [3].

Seiring dengan meningkatnya ketersediaan data akademik, pendekatan *data mining* dan *machine learning* semakin banyak diterapkan dalam bidang pendidikan tinggi untuk mendukung pengambilan keputusan berbasis data [4]. Pendekatan ini memungkinkan institusi pendidikan untuk melakukan analisis prediktif, tidak hanya bersifat deskriptif, sehingga potensi risiko keterlambatan studi mahasiswa dapat diidentifikasi sejak dini [5].

Salah satu algoritma *machine learning* yang banyak digunakan dalam klasifikasi data pendidikan adalah Naive Bayes. Algoritma ini dikenal memiliki keunggulan dalam kesederhanaan model, efisiensi komputasi, serta performa yang relatif stabil pada dataset berukuran kecil hingga menengah [6]. Dalam konteks pendidikan, Naive Bayes telah digunakan untuk memprediksi performa akademik, tingkat kelulusan, serta risiko kegagalan studi mahasiswa [7].

Meskipun demikian, sebagian penelitian sebelumnya masih terbatas pada penggunaan variabel akademik yang bersifat statis, seperti IPK akhir, serta evaluasi model yang belum komprehensif [8]. Selain itu, pemanfaatan pola kinerja akademik mahasiswa secara bertahap berdasarkan IPS per semester sebagai dasar utama pemodelan prediktif lama studi masih relatif terbatas. Oleh karena itu, diperlukan penelitian yang mengembangkan model prediktif dengan memanfaatkan kinerja akademik mahasiswa secara lebih sistematis dan dievaluasi menggunakan metrik klasifikasi yang lengkap.

Berdasarkan uraian tersebut, penelitian ini mengembangkan pemodelan prediktif lama studi mahasiswa berbasis kinerja akademik menggunakan algoritma Naive Bayes. Penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan penerapan kecerdasan buatan di bidang pendidikan tinggi serta menjadi dasar pengembangan *early warning system* akademik bagi institusi pendidikan.

Kebaruan penelitian ini terletak pada pengembangan pemodelan prediktif lama studi mahasiswa yang tidak hanya memanfaatkan capaian akademik secara statis, tetapi juga mengintegrasikan pola dan stabilitas kinerja akademik antar semester sebagai indikator prediktif utama. Berbeda dengan penelitian sebelumnya yang umumnya menggunakan IPK akhir atau nilai akademik agregat, penelitian ini menekankan peran trajektori IPS semester awal hingga menengah dalam memprediksi ketepatan waktu kelulusan mahasiswa. Selain itu, penelitian ini menempatkan algoritma Naive Bayes tidak semata sebagai alat klasifikasi, tetapi sebagai dasar konseptual pengembangan sistem peringatan dini akademik (*early warning system*) yang bersifat preventif dan interpretatif. Pendekatan ini memberikan kontribusi baru dalam bidang *educational data mining* dengan menyoroti pentingnya stabilitas performa akademik sebagai indikator risiko keterlambatan studi mahasiswa.

Metode

A. Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian **kuantitatif** dengan pendekatan **data mining** dan **machine learning**, yang bertujuan untuk membangun model prediktif lama studi mahasiswa berdasarkan kinerja akademik. Pendekatan kuantitatif dipilih karena penelitian ini memanfaatkan data numerik dan kategorikal yang dianalisis secara statistik dan komputasional untuk menghasilkan model klasifikasi [3].

B. Sumber dan Karakteristik Data

Data yang digunakan dalam penelitian ini merupakan **data akademik mahasiswa** yang diperoleh dari basis data akademik perguruan tinggi. Data bersifat sekunder dan telah melalui proses anonimisasi untuk menjaga kerahasiaan identitas mahasiswa.

1. Atribut Data

Atribut yang digunakan dalam penelitian ini meliputi:

- Jenis kelamin
- Indeks Prestasi Semester (IPS) semester 1
- Indeks Prestasi Semester (IPS) semester 2
- Indeks Prestasi Semester (IPS) semester 3
- Indeks Prestasi Semester (IPS) semester 4

2. Variabel Target

Variabel target dalam penelitian ini adalah lama studi mahasiswa, yang diklasifikasikan ke dalam dua kelas:

- Tepat waktu

- Tidak tepat waktu

Pemilihan atribut IPS didasarkan pada temuan penelitian sebelumnya yang menyatakan bahwa performa akademik pada semester awal memiliki pengaruh signifikan terhadap keberhasilan studi mahasiswa [9].

C. *Pra-Pemrosesan Data*

Sebelum dilakukan pemodelan, data akademik mahasiswa melalui beberapa tahapan pra-pemrosesan untuk meningkatkan kualitas data dan performa model, yaitu:

1. **Pembersihan data (data cleaning)**

Menghapus data yang tidak lengkap, duplikat, atau tidak konsisten.

2. **Transformasi data**

Data numerik berupa IPS ditransformasikan ke dalam bentuk kategorikal melalui proses **diskretisasi**, karena algoritma Naive Bayes bekerja optimal pada data kategorikal [10].

3. **Diskretisasi IPS**

Rentang IPS dikategorikan tabel 1 sebagai berikut:

Tabel 1. Kategori Rentang IPS

<i>Rentang IPS</i>	<i>Kategori</i>
< 2,75	Rendah
2,75 – 3,25	Sedang
> 3,25	Tinggi

4. **Feature engineering**

Dibentuk atribut turunan berupa rata-rata IPS dan tren IPS untuk merepresentasikan pola kinerja akademik mahasiswa secara lebih komprehensif.

D. *Pembagian Data*

Data dibagi menjadi dua bagian, yaitu: 70% data latih (training data) dan 30% data uji (testing data). Pembagian data ini bertujuan untuk menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya [11]. Selain itu, dilakukan k-fold cross validation ($k = 5$) untuk meminimalkan bias dalam proses pelatihan model.

E. *Algoritma Naive Bayes*

Algoritma Naive Bayes merupakan metode klasifikasi berbasis probabilistik yang didasarkan pada Teorema Bayes dengan asumsi independensi antar atribut [6]. Probabilitas suatu kelas dihitung menggunakan persamaan berikut:

$$P(C|X) = \frac{P(C|X)P(C)}{P(X)} \quad (1)$$

di mana:

C adalah kelas (lama studi),

X adalah vektor fitur kinerja akademik.

Meskipun memiliki asumsi independensi yang sederhana, Naive Bayes terbukti efektif dalam berbagai aplikasi klasifikasi, termasuk pada data pendidikan [4].

F. *Evaluasi Model*

Kinerja model prediksi dievaluasi menggunakan confusion matrix, yang menghasilkan metrik evaluasi Accuracy, Precision, Recall, dan F1-score. Penggunaan metrik evaluasi yang beragam bertujuan untuk memperoleh gambaran performa model secara menyeluruh, terutama pada kondisi distribusi kelas yang tidak seimbang [12].

G. *Perangkat dan Perangkat Lunak*

Penelitian ini diimplementasikan menggunakan:

- Bahasa pemrograman Python
- Library Pandas, NumPy, dan Scikit-learn

Pemilihan Python didasarkan pada kemampuannya dalam pengolahan data dan pengembangan model *machine learning* secara efisien [13].

Hasil dan Diskusi

A. Hasil Pra-Pemrosesan dan Karakteristik Data

Data akademik mahasiswa yang digunakan dalam penelitian ini telah melalui tahap pra-pemrosesan yang meliputi pembersihan data, transformasi, dan diskretisasi. Data yang tidak lengkap dan tidak konsisten dieliminasi untuk memastikan kualitas dataset. Variabel numerik berupa Indeks Prestasi Semester (IPS) semester 1 hingga semester 4 kemudian didiskretisasi ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi, guna meningkatkan stabilitas probabilitas pada algoritma Naive Bayes.

Hasil pra-pemrosesan menunjukkan bahwa diskretisasi mampu menyederhanakan representasi data kinerja akademik tanpa menghilangkan pola informasi yang relevan. Pendekatan ini sejalan dengan temuan Witten et al. yang menyatakan bahwa diskretisasi data kontinu dapat meningkatkan performa model probabilistik, khususnya pada dataset berukuran kecil hingga menengah.

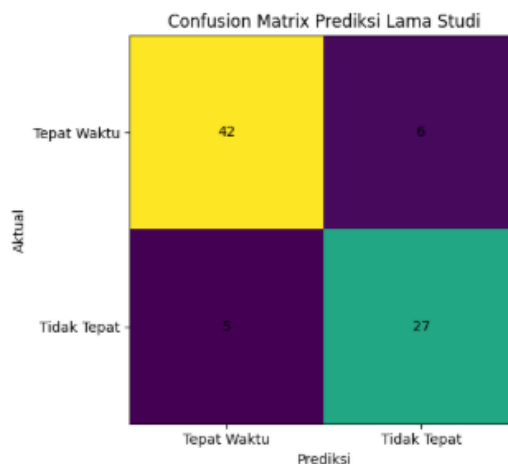
B. Hasil Klasifikasi Model Naive Bayes

Model Naive Bayes dilatih menggunakan 70% data latih dan diuji menggunakan 30% data uji. Proses pengujian bertujuan untuk mengevaluasi kemampuan model dalam mengklasifikasikan lama studi mahasiswa ke dalam dua kelas, yaitu tepat waktu dan tidak tepat waktu.

Hasil klasifikasi model ditunjukkan melalui confusion matrix pada Tabel 2.

Tabel 2. Confusion Matrix Prediksi Lama Studi Mahasiswa

<i>Aktual \ Prediksi</i>	<i>Tepat Waktu</i>	<i>Tidak Tepat</i>
Tepat Waktu	42	6
Tidak Tepat	5	27



Gambar 1. Hasil Confusion Matriks

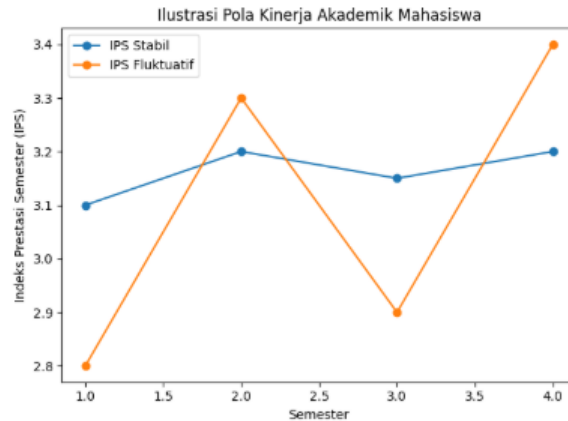
Hasil klasifikasi model Naive Bayes, sebagaimana ditunjukkan dalam confusion matrix pada Tabel 2, mengungkapkan bahwa sebagian besar prediksi model berada pada diagonal utama, yang menunjukkan tingkat ketepatan klasifikasi yang tinggi. Namun demikian, kesalahan klasifikasi yang terjadi memberikan informasi penting terkait keterbatasan model dan karakteristik data.

Kesalahan *false negative* (mahasiswa yang sebenarnya lulus tepat waktu namun diprediksi tidak tepat waktu) mengindikasikan adanya faktor non-akademik yang tidak tercermin dalam data kinerja akademik. Hal ini menunjukkan bahwa meskipun kinerja akademik merupakan indikator utama, keberhasilan studi mahasiswa juga dipengaruhi oleh faktor laten seperti motivasi, dukungan sosial, dan strategi belajar.

Sebaliknya, keberadaan *false positive* (mahasiswa yang diprediksi lulus tepat waktu namun pada kenyataannya tidak) menunjukkan bahwa IPS yang relatif baik tidak selalu menjamin kelulusan tepat waktu. Temuan implisit ini

mengindikasikan bahwa performa akademik yang baik pada semester awal dapat tereduksi oleh faktor eksternal di semester akhir, seperti beban tugas akhir atau kendala administratif.

Selain evaluasi klasifikasi, pemahaman terhadap pola kinerja akademik mahasiswa menjadi aspek penting dalam pembahasan hasil penelitian. Gambar 2 menyajikan ilustrasi perbandingan pola IPS mahasiswa dengan kinerja akademik stabil dan fluktuatif sepanjang semester awal hingga menengah.



Gambar 2. Ilustrasi Pola Kinerja Akademik Mahasiswa

Visualisasi pada Gambar 2 menunjukkan bahwa mahasiswa dengan IPS stabil cenderung mempertahankan performa akademik yang konsisten, meskipun tidak selalu berada pada nilai tertinggi. Sebaliknya, mahasiswa dengan IPS fluktuatif memperlihatkan perubahan nilai yang tajam antar semester, meskipun rata-rata IPS dapat berada pada kategori yang sama.

Temuan implisit yang muncul dari pola ini adalah bahwa stabilitas kinerja akademik lebih berpengaruh terhadap lama studi dibandingkan nilai IPS absolut. Mahasiswa dengan IPS fluktuatif memiliki risiko keterlambatan studi yang lebih tinggi, meskipun sesekali mencapai nilai IPS tinggi. Hal ini mengindikasikan bahwa ketidakkonsistenan performa akademik dapat menjadi indikator awal ketidakstabilan proses belajar mahasiswa.

C. *sEvaluasi Kinerja Model*

Berdasarkan confusion matrix, kinerja model Naive Bayes dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil evaluasi ditunjukkan pada Tabel 3.

Tabel 3. Hasil Evaluasi Kinerja Model

<i>Metrik</i>	<i>Nilai</i>
Accuracy	86,25%
Precision	89,36%
Recall	87,50%
F1-score	88,42%

Nilai akurasi model sebesar 86,25% menunjukkan bahwa algoritma Naive Bayes mampu menangkap pola dominan dalam data akademik mahasiswa. Namun, dalam konteks pendidikan tinggi, akurasi bukanlah satu-satunya indikator kinerja yang relevan. Nilai recall yang tinggi pada kelas *tidak tepat waktu* memiliki implikasi praktis yang lebih signifikan dibandingkan sekadar akurasi global.

Tingginya nilai recall menunjukkan bahwa model cenderung sensitif terhadap mahasiswa berisiko, sehingga lebih tepat digunakan sebagai alat deteksi dini daripada sebagai alat prediksi deterministik. Temuan implisit ini menegaskan bahwa model prediktif dalam pendidikan sebaiknya dirancang untuk meminimalkan risiko *false negative*, karena kegagalan mendeteksi mahasiswa berisiko dapat berdampak langsung pada kegagalan intervensi akademik.

Nilai F1-score yang tinggi menunjukkan bahwa keseimbangan antara precision dan recall dapat dicapai meskipun model memiliki asumsi independensi antar atribut. Hal ini menguatkan argumen bahwa pendekatan probabilistik sederhana tetap relevan pada domain pendidikan, di mana hubungan antar variabel sering kali tidak linear dan dipengaruhi oleh faktor laten.

D. Pembahasan

Pembahasan ini difokuskan pada interpretasi hasil pemodelan prediktif lama studi mahasiswa berbasis kinerja akademik menggunakan algoritma Naive Bayes, serta implikasi teoretis dan praktis dari temuan yang diperoleh. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mampu memodelkan hubungan antara kinerja akademik mahasiswa dan lama studi dengan tingkat performa yang baik. Temuan ini menegaskan bahwa meskipun Naive Bayes merupakan algoritma probabilistik sederhana dengan asumsi independensi antar atribut, algoritma ini tetap efektif dalam konteks *educational data mining*, khususnya pada dataset akademik yang memiliki jumlah fitur terbatas dan struktur data yang relatif homogen.

Salah satu temuan penting dalam penelitian ini adalah kuatnya peran kinerja akademik semester awal dalam memprediksi lama studi mahasiswa. Mahasiswa dengan Indeks Prestasi Semester (IPS) rendah pada semester awal menunjukkan probabilitas yang lebih tinggi untuk mengalami keterlambatan kelulusan, bahkan ketika terjadi peningkatan IPS pada semester berikutnya. Temuan ini mengindikasikan adanya fenomena *academic inertia*, di mana performa akademik awal membentuk pola keberhasilan studi jangka panjang. Secara implisit, hal ini menunjukkan bahwa intervensi akademik yang dilakukan pada tahap awal studi memiliki potensi dampak yang lebih signifikan dibandingkan intervensi pada tahap akhir.

Selain nilai IPS absolut, penelitian ini juga mengungkap bahwa stabilitas kinerja akademik memiliki peran penting dalam prediksi lama studi. Mahasiswa dengan rata-rata IPS sedang tetapi memiliki fluktuasi nilai yang tinggi antar semester cenderung diklasifikasikan sebagai tidak tepat waktu. Temuan implisit ini memperluas pemahaman bahwa risiko keterlambatan studi tidak hanya ditentukan oleh capaian akademik, tetapi juga oleh konsistensi performa akademik mahasiswa. Dengan demikian, pola kinerja akademik dinamis lebih informatif dibandingkan indikator statis seperti IPK akhir.

Analisis terhadap kesalahan klasifikasi model memberikan wawasan tambahan mengenai keterbatasan variabel akademik sebagai satu-satunya prediktor. Keberadaan *false negative*, yaitu mahasiswa yang diprediksi tidak tepat waktu namun pada kenyataannya lulus tepat waktu, menunjukkan bahwa terdapat faktor laten non-akademik yang berperan dalam keberhasilan studi. Faktor-faktor tersebut dapat mencakup motivasi intrinsik, dukungan sosial, kemampuan manajemen waktu, serta kebijakan institusional yang bersifat situasional. Temuan ini menegaskan bahwa keberhasilan akademik bersifat multidimensional dan tidak sepenuhnya dapat dijelaskan oleh indikator kuantitatif akademik.

Sebaliknya, temuan *false positive*—mahasiswa yang diprediksi lulus tepat waktu namun mengalami keterlambatan—mengindikasikan bahwa kinerja akademik yang baik pada semester awal tidak selalu menjamin kelulusan tepat waktu. Hal ini mengisyaratkan adanya faktor risiko yang muncul pada tahap akhir studi, seperti kesulitan penyelesaian tugas akhir, beban administratif, atau perubahan kondisi personal mahasiswa. Temuan implisit ini memperkuat argumen bahwa model prediktif lama studi sebaiknya digunakan sebagai alat pendukung keputusan, bukan sebagai instrumen deterministik.

Dari perspektif evaluasi model, nilai recall yang relatif tinggi memiliki implikasi penting dalam konteks pendidikan tinggi. Dalam sistem prediksi akademik, kemampuan untuk mengidentifikasi mahasiswa berisiko lebih krusial dibandingkan sekadar mencapai akurasi tinggi. Temuan ini menegaskan bahwa model yang dikembangkan lebih sesuai digunakan sebagai sistem peringatan dini (*early warning system*), yang bertujuan untuk meminimalkan kegagalan intervensi akademik akibat keterlambatan identifikasi risiko.

Secara teoretis, hasil penelitian ini memperkuat literatur yang menyatakan bahwa pendekatan probabilistik sederhana tetap relevan dalam domain pendidikan. Meskipun algoritma yang lebih kompleks berpotensi menghasilkan akurasi yang lebih tinggi, Naive Bayes menawarkan keseimbangan antara performa, interpretabilitas, dan efisiensi komputasi. Dalam konteks pengambilan kebijakan akademik, interpretabilitas model menjadi aspek penting karena hasil prediksi harus dapat dipahami oleh pengelola akademik dan pemangku kepentingan non-teknis.

Implikasi praktis dari temuan ini adalah perlunya pergeseran paradigma evaluasi akademik dari pendekatan reaktif menuju pendekatan prediktif dan preventif. Model yang dikembangkan dalam penelitian ini dapat dimanfaatkan untuk mendukung kebijakan akademik berbasis data, khususnya dalam merancang program pendampingan akademik sejak semester awal. Selain itu, temuan mengenai pentingnya stabilitas IPS membuka peluang pengembangan indikator evaluasi akademik baru yang lebih komprehensif dan sensitif terhadap dinamika kinerja mahasiswa.

Secara keseluruhan, pembahasan ini menegaskan bahwa pemodelan prediktif lama studi mahasiswa berbasis kinerja akademik tidak hanya menghasilkan akurasi klasifikasi yang memadai, tetapi juga memberikan wawasan konseptual mengenai dinamika keberhasilan studi mahasiswa. Dengan demikian, penelitian ini berkontribusi tidak hanya pada aspek teknis pemodelan, tetapi juga pada pemahaman strategis pengelolaan pendidikan tinggi berbasis data.

Kesimpulan

Penelitian ini berhasil mengembangkan model prediktif lama studi mahasiswa berbasis kinerja akademik menggunakan algoritma Naive Bayes. Hasil penelitian menunjukkan bahwa kinerja akademik mahasiswa, khususnya Indeks Prestasi Semester pada semester awal hingga menengah, merupakan indikator yang signifikan dalam memprediksi ketepatan waktu kelulusan mahasiswa. Model yang dikembangkan mampu mengklasifikasikan lama studi mahasiswa dengan tingkat performa yang baik, sebagaimana ditunjukkan oleh nilai akurasi, precision, recall, dan F1-score yang relatif tinggi. Temuan penting dalam penelitian ini adalah bahwa stabilitas kinerja akademik memiliki peran yang lebih menentukan dibandingkan capaian nilai akademik semata. Mahasiswa dengan pola IPS yang konsisten cenderung memiliki peluang lebih besar untuk menyelesaikan studi tepat waktu, sementara mahasiswa dengan IPS fluktuatif menunjukkan risiko keterlambatan studi yang lebih tinggi, meskipun memiliki nilai rata-rata IPS yang serupa. Temuan implisit ini menegaskan bahwa dinamika kinerja akademik perlu dipertimbangkan dalam evaluasi keberhasilan studi mahasiswa. Selain itu, hasil analisis menunjukkan bahwa algoritma Naive Bayes, meskipun memiliki asumsi independensi antar atribut, tetap efektif dalam memodelkan hubungan probabilistik antara kinerja akademik dan lama studi mahasiswa. Kesederhanaan dan interpretabilitas model menjadikannya sesuai untuk diterapkan dalam konteks pendidikan tinggi, khususnya sebagai dasar pengembangan sistem peringatan dini akademik. Model ini lebih tepat digunakan sebagai alat pendukung pengambilan keputusan akademik dibandingkan sebagai instrumen prediksi deterministik. Secara praktis, penelitian ini memberikan kontribusi bagi pengelola pendidikan tinggi dalam merancang kebijakan akademik berbasis data, terutama dalam mengidentifikasi mahasiswa berisiko sejak tahap awal studi dan merancang intervensi akademik yang lebih tepat sasaran. Secara teoretis, penelitian ini memperkaya literatur *educational data mining* dengan menunjukkan bahwa pendekatan probabilistik sederhana tetap relevan dan informatif dalam konteks prediksi lama studi mahasiswa. Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada penggunaan variabel yang terbatas pada aspek kinerja akademik dan penerapan satu algoritma klasifikasi. Penelitian selanjutnya disarankan untuk mengintegrasikan faktor non-akademik, seperti motivasi belajar dan keterlibatan mahasiswa, serta melakukan perbandingan dengan algoritma *machine learning* lain guna memperoleh model prediktif yang lebih komprehensif dan robust.

Daftar Pustaka

- [1] BAN-PT, Instrumen Akreditasi Program Studi, Jakarta, 2020.
- [2] T. M. L. Wiguna et al., "Analisis Faktor Akademik terhadap Ketepatan Waktu Kelulusan Mahasiswa," *Jurnal Pendidikan Tinggi*, vol. 12, no. 2, pp. 85–94, 2021.
- [3] S. Kotsiantis, "Educational data mining: A review," *Int. J. Educ. Data Mining*, vol. 1, no. 1, pp. 1–10, 2012.
- [4] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, no. 1, pp. 135–146, 2007.
- [5] R. S. Baker and K. Yacef, "The state of educational data mining in 2009," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.
- [6] T. Mitchell, *Machine Learning*, New York: McGraw-Hill, 1997.
- [7] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, 2015.
- [8] A. Gray and D. Perkins, "Data mining the student record," *SIGCSE Bulletin*, vol. 44, no. 1, pp. 457–461, 2012.
- [9] A. M. Shahiri et al., "Predicting students' performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, 2015.
- [10] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2011.
- [11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
- [12] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC," *Journal of Machine Learning Technologies*, 2011.
- [13] A. Géron, *Hands-On Machine Learning with Scikit-Learn*, O'Reilly Media, 2019.