



Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma Naïve Bayes Berbasis Particle Swarm Optimization

Hadi Permana

Universitas Budi Luhur, Indonesia
Email: hadipermana78@gmail.com

Keywords:

Naïve Bayes; Particle Swarm Optimization; Data Mining

Abstract

The success of students in completing their studies within the target timeframe is a crucial indicator in the accreditation assessment of a study program at the university level. Therefore, the ability to estimate the probability of on-time graduation is essential for study program administrators as a basis for policy-making. This study aims to develop a model for forecasting on-time graduation among students in the Architecture Study Program at Trisakti University. The method applied is the Naïve Bayes algorithm enhanced with Particle Swarm Optimization (PSO). The variables used encompass two dimensions: academic and social. Based on validation using the k-fold cross-validation method, the standalone Naïve Bayes algorithm produced an accuracy rate of 79.22% with an AUC value of 0.859. On the other hand, the application of PSO optimization successfully boosted performance to an accuracy of 85.71% and an AUC value of 0.879. Based on these results, it can be concluded that the application of the Naïve Bayes algorithm optimized with Particle Swarm Optimization is effective in improving the accuracy of predicting the graduation of students on time. The resulting model has the potential to be used as a decision support system in early student academic monitoring.

Kata Kunci:

Naïve Bayes; Particle Swarm Optimization; Data Mining

Abstrak

Keberhasilan mahasiswa menyelesaikan studi sesuai target waktu menjadi indikator krusial dalam penilaian akreditasi suatu program studi di tingkat perguruan tinggi. Oleh karena itu, kemampuan untuk memperkirakan probabilitas kelulusan tepat waktu sangat dibutuhkan oleh pengelola program studi sebagai dasar pengambilan kebijakan. Studi ini bertujuan untuk membangun model peramalan kelulusan tepat waktu pada mahasiswa Program Studi Arsitektur Universitas Trisakti. Metode yang diaplikasikan adalah Algoritma Naïve Bayes yang diperkuat dengan optimasi Particel Swarm Optimization (PSO). Variabel yang dipakai mencakup dua dimensi, yakni akademik dan sosial. Berdasarkan validasi memakai metode k-fold cross validation, Algoritma Naïve Bayes standalone menghasilkan tingkat akurasi 79,22% dengan nilai AUC 0,859. Di sisi lain, penerapan optimasi PSO sukses mendongkrak performa hingga mencapai akurasi 85,71% dan nilai AUC 0,879. Berdasarkan hasil tersebut, dapat disimpulkan bahwa penerapan algoritma Naïve Bayes yang dioptimasi dengan Particle Swarm Optimization efektif dalam meningkatkan akurasi prediksi kelulusan mahasiswa tepat waktu. Model yang dihasilkan memiliki potensi untuk digunakan sebagai sistem pendukung keputusan dalam pemantauan akademik mahasiswa secara dini.

PENDAHULUAN

Perkembangan teknologi informasi saat ini menjadi kunci dalam menghadirkan data yang valid guna mendukung kemajuan suatu lembaga pendidikan. Dalam konteks ini, informasi mengenai performa akademik mahasiswa menjadi dasar utama dalam perumusan kebijakan (Asif et al., 2017; Burgos et al., 2018; Costa et al., 2017; Masci et al., 2018). Akan tetapi, ketersediaan data mentah yang melimpah di database sering kali belum bisa langsung dimanfaatkan, sehingga diperlukan pengolahan lebih lanjut untuk menghasilkan informasi yang tepat sasaran (Bernacki et al., 2020; Fernandes et al., 2019; Rebai et al., 2020; Rizvi et al., 2019).

Universitas Trisakti sendiri telah mengimplementasikan Student Information System (SIS) sebagai tempat mengarsipkan seluruh data dan nilai mahasiswa. Meski demikian, data tersebut baru sebatas rekaman statis dan belum dimanfaatkan secara maksimal menjadi *insight* yang mendukung pengambilan keputusan. Padahal, salah satu tolak ukur vital dalam akreditasi perguruan tinggi adalah tingkat kelulusan tepat waktu beserta rata-rata IPK lulusan. Ketiadaan sistem peramalan kelulusan di Program Studi Arsitektur membuat manajemen kesulitan menyusun strategi pemantauan kinerja mahasiswa secara dini.

Teknik *data mining* hadir sebagai solusi untuk mengekstrak pola tersembunyi dari database guna mendukung prediksi tersebut. Sebagaimana dijelaskan oleh Turban et al., (2007), *data mining* merupakan proses ekstraksi pengetahuan yang memadukan statistika, matematika, hingga *machine learning* dari sekumpulan data besar.

Penelitian ini memilih Algoritma Naïve Bayes setelah melakukan uji coba awal menggunakan software RapidMiner pada data prodi Arsitektur, di mana metode ini mencatat akurasi 80,44%, lebih unggul dibanding algoritma C4.5 yang hanya 75,64%. Untuk memaksimalkan potensi akurasi, algoritma Naïve Bayes selanjutnya dikombinasikan dengan Particle Swarm Optimization (PSO) dengan harapan mampu menghasilkan model prediksi pola kelulusan mahasiswa yang lebih *robust* dan akurat.

Berbagai studi terdahulu telah mengeksplorasi prediksi kelulusan menggunakan berbagai metode. Kamagi & Hansun, (2014) menerapkan algoritma C4.5 dengan parameter IPS hingga semester enam dan berhasil mendapatkan akurasi 87,5%. Berbeda dengan Gian (2014) yang menemukan bahwa C4.5 (77,354%) sedikit lebih unggul dibanding Naive Bayes, (74,094%) dalam prediksi ketepatan waktu kelulusan.

Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan Naive Bayes dan Algoritma C4.5 untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa (GIAN, 2014). Komparasi antara naïve bayes dan algoritma C 4.5 dengan mengukur kinerja kedua metode tersebut menggunakan metode pengujian confusion matrix, kemudian diketahui bahwa C4.5 memiliki nilai akurasi yang lebih baik yakni sebesar 77,354% , sedangkan naive bayes memiliki nilai akurasi mencapai 74,094%. Kemudian berdasarkan perbandingan kinerja kedua metode tersebut, metode dengan pencapaian nilai akurasi terbaik akan diimplementasikan dalam bentuk sebuah *Decision Support System* (Chung & Lee, 2019; Cruz-Jesus et al., 2020; Waheed et al., 2020).

Supriyanti et al., (2016) melakukan studi komparatif antara algoritma C4.5 dan Naive Bayes dalam menentukan ketepatan pilihan konsentrasi mahasiswa. Penelitian mereka menunjukkan bahwa penerapan teknik *forward selection* sebagai metode seleksi fitur mampu mengangkat performa kedua algoritma. Sebelum diberikan seleksi fitur, akurasi C4.5 tercatat

sebesar 84,43% dan Naive Bayes 78,47%. Setelah *forward selection* diterapkan, nilai tersebut mengalami kenaikan masing-masing menjadi 84,98% dan 82,01%. Dari hasil perbandingan akhir, algoritma C4.5 yang telah dipadukan dengan seleksi fitur terbukti menjadi metode terbaik dengan raihan akurasi paling tinggi, yakni 84,98%.

Studi yang dilakukan oleh Arga et al., (2017) di Jurusan Teknik Informatika IST Akprind Yogyakarta memanfaatkan algoritma C4.5 untuk membangun model peramalan masa studi mahasiswa. Variabel prediktor yang diinputkan meliputi asal sekolah, mata kuliah konsentrasi yang dipilih, IPK saat semester 7, serta jumlah SKS yang telah ditempuh hingga semester tersebut. Melalui sistem yang dikembangkan, pihak program studi diharapkan bisa mengidentifikasi faktor-faktor dominan yang memengaruhi lama studi sekaligus memperkirakan waktu kelulusan mahasiswa. Berdasarkan evaluasi terhadap aplikasi yang dibuat, algoritma C4.5 berhasil mengklasifikasikan mahasiswa ke dalam kelompok masa studi yang sudah ditentukan dengan tingkat akurasi mencapai 72,25% dan tingkat kesalahan (*error*) sebesar 27,75%.

Muhamad beserta tim peneliti (2017) mengkaji pendekatan untuk mengoptimalkan kinerja Naïve Bayes Classifier menggunakan dataset Iris. Di satu sisi, metode Naïve Bayes dikenal memiliki sejumlah keunggulan seperti proses komputasi yang cepat, algoritma yang tidak rumit, dan tingkat akurasi yang baik. Namun di sisi lain, metode ini memiliki kelemahan, yakni ketidakmampuan nilai probabilitasnya untuk mengukur tingkat keyakinan akurasi suatu prediksi secara pasti, performanya yang acapkali di bawah algoritma C4.5, serta kurangnya efektivitas dalam hal pemilihan atribut. Untuk mengatasi berbagai kelemahan tersebut, penelitian ini mengusulkan penerapan algoritma Particle Swarm Optimization (PSO) yang berfungsi untuk melakukan pembobotan pada atribut, sehingga mampu mendongkrak tingkat akurasi dari klasifikasi Naïve Bayes itu sendiri.

Meskipun sejumlah penelitian tersebut menunjukkan bahwa metode klasifikasi mampu digunakan untuk memprediksi kelulusan mahasiswa, masih terdapat beberapa celah penelitian. Sebagian besar penelitian terdahulu lebih banyak menggunakan algoritma C4.5 atau hanya membandingkan kinerja beberapa algoritma tanpa mengembangkan model optimasi yang lebih akurat. Selain itu, variabel yang digunakan dalam prediksi umumnya lebih menitikberatkan pada aspek akademik, seperti IPK, IPS, dan jumlah SKS, sedangkan aspek sosial mahasiswa masih relatif jarang dilibatkan sebagai faktor prediktor. Padahal, faktor sosial seperti asal daerah, pekerjaan orang tua, jalur penerimaan, dan riwayat cuti juga berpotensi memengaruhi proses studi dan ketepatan waktu kelulusan mahasiswa.

Berdasarkan kondisi tersebut, kebaruan (*novelty*) dalam penelitian ini terletak pada pengembangan model prediksi kelulusan mahasiswa tepat waktu dengan mengombinasikan algoritma Naïve Bayes dan Particle Swarm Optimization (PSO), serta mengintegrasikan dua dimensi variabel sekaligus, yaitu aspek akademik dan aspek sosial. Penelitian ini tidak hanya berfokus pada peningkatan akurasi model melalui optimasi algoritma, tetapi juga menyoroti pentingnya pelibatan atribut sosial untuk memperkaya informasi prediktif dalam konteks pendidikan tinggi. Dengan demikian, penelitian ini diharapkan mampu menghasilkan model prediksi yang lebih akurat, komprehensif, dan relevan bagi kebutuhan pengelolaan akademik di Program Studi Arsitektur Universitas Trisakti.

Tujuan penelitian ini adalah untuk membangun model prediksi kelulusan mahasiswa

tepat waktu menggunakan algoritma Naïve Bayes berbasis Particle Swarm Optimization, serta menganalisis pengaruh penambahan atribut sosial terhadap peningkatan akurasi model prediksi. Penelitian ini juga bertujuan untuk membandingkan performa algoritma Naïve Bayes sebelum dan sesudah optimasi PSO, sehingga dapat diketahui efektivitas optimasi dalam menghasilkan model klasifikasi yang lebih baik. Penelitian ini diharapkan memberikan manfaat secara teoritis maupun praktis. Secara teoritis, penelitian ini dapat memperkaya kajian di bidang data mining pendidikan, khususnya dalam pengembangan model prediksi kelulusan mahasiswa dengan pendekatan klasifikasi berbasis optimasi. Penelitian ini juga memberikan kontribusi pada pemahaman bahwa prediksi akademik tidak hanya dipengaruhi oleh variabel akademik, tetapi juga oleh faktor sosial mahasiswa. Secara praktis, hasil penelitian ini diharapkan dapat menjadi referensi bagi pengelola program studi dalam mengidentifikasi mahasiswa yang berpotensi terlambat lulus sejak dini, sehingga langkah pendampingan akademik dapat dilakukan secara lebih terarah. Selain itu, model yang dihasilkan juga dapat digunakan sebagai dasar pengembangan sistem pendukung keputusan dalam pemantauan kinerja mahasiswa dan perumusan kebijakan akademik di perguruan tinggi.

METODE PENELITIAN

Metodologi Penelitian

Tahap CRISP-DM (*Cross Standard Industries Process for Data Mining*). Berdasarkan arsitektur *data mining* CRISP-DM (*Cross Standart Industries Process for Data Mining*), maka dilakukan tahap-tahap *data mining* sebagai berikut:

1. Business Understanding

Adanya kesulitan untuk memprediksi kelulusan mahasiswa tepat waktu pada prodi Arsitektur Universitas Trisakti. Untuk itu dalam penelitian ini algoritma Naïve Bayes akan dioptimasi dengan menggunakan *Particle Swarm Optimization* untuk pembuatan model prediksi kelulusan mahasiswa tepat waktu pada Prodi Arsitektur.

2. Data Understanding

Untuk pemilihan prediksi kelulusan mahasiswa tepat waktu, didapat data dari Prodi Arsitektur Universitas Trisakti, yang terdiri dari 11 atribut. Atribut-atribut yang menjadi parameter terlihat pada tabel III-1 yaitu:

Tabel 1. Atribut dan Nilai Kategori

NO	Atribut	Nilai	Keterangan
1	Jenis Kelamin	P	Jenis Kelamin
		W	
2	IPS Sem 1		Indek Prestasi Semester 1
3	IPS Sem 2		Indek Prestasi Semester 2
4	IPS Sem 3		Indek Prestasi Semester 3
5	IPS Sem 4		Indek Prestasi Semester 4
6	IPK		Indek Prestasi Kumulatif
7	Total SKS Sem 4		Total SKS yang telah lulus hingga semester 4
8	Cuti	Cuti	Cuti yang pernah diambil

		Tidak	
9	Jalur Penerimaan		Jalur masuk dari mahasiswa
		Reguler	
		PSSB	
10	Pekerjaan Orang Tua		Pekerjaan Orang Tua
		PNS	
		PS	
		TNI/POLRI	
		Wiraswasta	
11	Asal Daerah		Asal Daerah dari mahasiswa
		Jakarta	
		Luar Jakarta	
Sumber : Dokumen Pribadi			

Pemilihan atribut tersebut dibagi menjadi dua faktor yaitu aspek akademik seperti : IPS1, IPS2, IPS3, IPS4, IPK, Jumlah SKS dan aspek sosial antara lain : Jenis Kelamin, Asal Daerah, Pekerjaan Orang Tua, Jalur Penerimaan, Cuti. Aspek sosial budaya adalah faktor yang ditambahkan penulis untuk mengetahui sejauh mana atribut-atribut tersebut turut mempengaruhi tingkat keakurasian hasil prediksi kelulusan mahasiswa tepat waktu.

3. Data Preparation

Data yang diperoleh untuk penelitian ini sebanyak 300 *record* mahasiswa. Untuk mendapatkan data yang baik, ada beberapa teknik *preprocessing* yang dapat dilakukan untuk mendapatkan sampel yang baik. Pada penelitian ini penulis menggunakan rumus slovin untuk mencari jumlah sampel yang sesuai dengan jumlah data yang ada dan hasilnya sebanyak 154 *record* dialokasikan untuk *training data* dan 100 *record* dialokasikan untuk *testing data*.

4. Modeling Fase

Tahap ini juga disebut tahap learning karena pada tahap ini data training diklasifikasi oleh model kemudian menghasilkan sejumlah aturan. Pada penelitian ini, pembuatan model menggunakan algoritma Naïve Bayes dan *Particle Swarm Optimization*.

5. Evaluation Fase

Pada tahap evaluation dilakukan pengujian terhadap pola yang sudah terbentuk guna mendapatkan informasi yang akurat. Evaluasi dan validasi menggunakan *precision*, *recall* dan akurasi.

6. Deployment

Setelah pembentukan model dan dilakukan analisa serta pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini juga diterapkan model yang paling akurat untuk penentuan klasifikasi prediksi kelulusan mahasiswa tepat waktu.

Confusion Matrix dan Kurva ROC

Metode *confusion matrix* dan kurva ROC merepresentasikan hasil evaluasi model dengan menggunakan tabel matriks, jika *dataset* terdiri dari dua kelas, kelas pertama dianggap positif dan kelas kedua dianggap negative. Evaluasi menggunakan confusion matrix menghasilkan

nilai akurasi, *precision*, *recall*. Akurasi dalam klasifikasi merupakan persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. *Precision* merupakan proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* merupakan proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar.

Tabel 2. Confusion Matrix

Correct Classification	Classified as	
	+	-
+	True Positive (TP)	False Negative (FN)
-	False Positive (FP)	True Negative (TN)

Sumber: Hasil pengolahan data menggunakan model klasifikasi

HASIL DAN PEMBAHASAN

Pengelompokkan Data

Data yang digunakan berasal dari SIS (*student information system*) Jurusan Arsitektur Universitas Trisakti, adapun record mahasiswa yang digunakan untuk data testing sebanyak 154 dan atribut yang digunakan sebanyak 12 atribut, dimana 11 atribut predictor dan 1 atribut hasil. Dari 12 atribut tersebut terdapat atribut dengan aspek sosial seperti pekerjaan orang tua, cuti, asal daerah dan jalur penerimaan, sisanya atribut dengan aspek akademik. Pengujian dilakukan dengan menggunakan seluruh data training dengan *confusion matrix* yang terdiri dari *accuracy*, *precision* dan *recall* dilakukan pada dataset yang diolah dengan menggunakan algoritma Naïve Bayes berbasis Particle Swarm Optimization.

Evaluasi

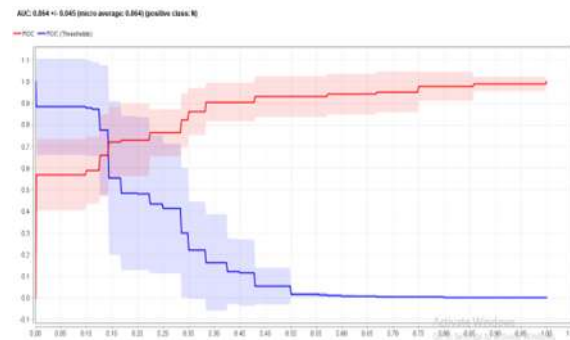
Pengujian dilakukan dengan menggunakan Confusion Matrix hasil dari, pengujian naïve bayes dapat dilihat pada tabel di bawah ini

Tabel 3. Confusion Matrix dengan Naïve Bayes

Confusion Matrix		
	N	Y
N	62	18
Y	14	60

Sumber: Hasil pengolahan data penelitian menggunakan algoritma Naïve Bayes

Kurva ROC Naïve Bayes



Gambar 1. Nilai AUC Algoritma Naïve Bayes

Sumber: Hasil pengolahan data penelitian menggunakan algoritma Naïve Bayes berbasis Particle Swarm Optimization

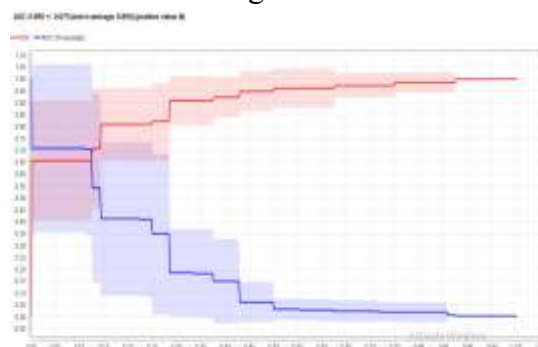
Tabel confusion matrix Naïve Bayes dan PSO dapat dilihat pada tabel di bawah ini.

Tabel 4. Confusion Matrix dengan Naïve Bayes berbasis PSO

Confusion Matrix		
	N	Y
N	67	13
Y	9	65

Sumber: Hasil pengolahan data penelitian menggunakan algoritma Naïve Bayes berbasis Particle Swarm Optimization

Kurva AUC Naïve Bayes berbasis PSO sebagai berikut:



Gambar 2. Nilai AUC algoritma Naïve Bayes Berbasis PSO

Sumber: Hasil pengolahan data penelitian menggunakan algoritma Naïve Bayes berbasis PSO (output RapidMiner)

Pengujian confusion matrix untuk dataset yang diolah menggunakan algoritma Naïve Bayes memiliki nilai akurasi 79,22% dan nilai AUC sebesar 0,864, sedangkan untuk Algoritma Naive Bayes berbasis PSO memiliki tingkat akurasi yaitu 85,71% dan nilai AUC 0,899 yang termasuk kategori ‘good classification’.

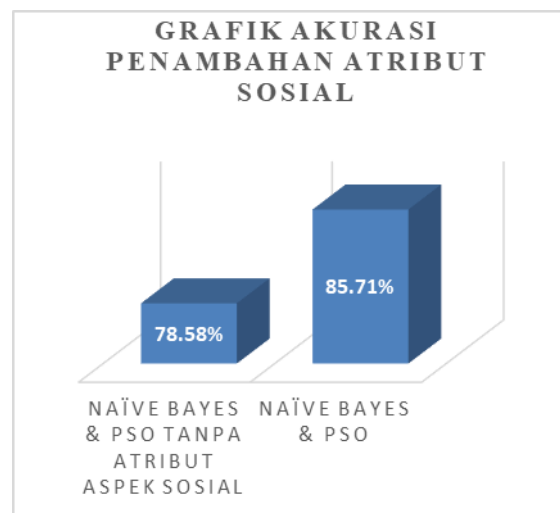
Tabel 5. Perbandingan Tingkat Akurasi, AUC algoritma Naïve Bayes dan Naïve Bayes Berbasis PSO

Prediksi	Naïve Bayes	Naïve Bayes + PSO
Prediksi tepat waktu	87%	91%
Prediksi tidak tepat waktu	72%	74%
Tingkat Akurasi	79,22%	85,71%
AUC	0,864	0,899

Sumber: Hasil analisis perbandingan kinerja model klasifikasi, diolah penulis

Atribut Aspek Sosial dan Atribut Aspek Akademik

Aspek sosial pada atribut yang terdapat pada data set ternyata sangat berpengaruh pada penelitian ini, pada pengujian keakurasian dari Naïve Bayes dan PSO didapat hasil sebagai berikut, hasil dari pelibatan fitur tentang aspek sosial (meliputi asal daerah, pekerjaan orang tua, jalur penerimaan) mampu menghasilkan akurasi sebesar 85,71% sedangkan tanpa aspek sosial tersebut, tingkat akurasi hanya sebesar 78,58 %. Dapat disimpulkan bahwa penambahan aspek sosial pada atribut data latih terbukti meningkatkan tingkat akurasi dari Naïve Bayes dan optimasi PSO.



Gambar 3. Grafik Penambahan Atribut Sosial

Sumber: Hasil analisis data penelitian, diolah penulis

KESIMPULAN

Metode klasifikasi data mining sangat baik untuk diimplementasikan ke dalam prototipe sistem informasi prediksi kelulusan mahasiswa tepat waktu. Dengan penambahan faktor sosial mampu meningkatkan tingkat keakurasian dengan hasil 85.71% dan untuk atribut tanpa penambahan faktor sosial memiliki tingkat akurasi 78.58%, terbukti faktor sosial berpengaruh untuk menambah tingkat keakurasian dari Naïve Bayes. Dapat disimpulkan bahwa *Particle Swarm Optimization* dapat meningkatkan Akurasi dari Algoritma Naïve Bayes. Hal ini dapat dilihat bahwa algoritma Naïve Bayes berbasis PSO memiliki nilai akurasi sebesar 85.71% dengan nilai AUC 0,899 yang termasuk kategori *good classification*. Sedangkan untuk Algoritma Naïve Bayes tanpa PSO memiliki Akurasi 79,22% dengan nilai AUC 0,864.

Adapun saran yang dapat diberikan berdasarkan hasil penelitian ini adalah sebagai berikut. Pertama, bagi pihak program studi, model prediksi ini dapat diimplementasikan

sebagai sistem pendukung keputusan untuk melakukan pemantauan dini terhadap potensi keterlambatan kelulusan mahasiswa, sehingga intervensi akademik dapat dilakukan secara lebih tepat sasaran. Kedua, penelitian selanjutnya disarankan untuk menambahkan variabel lain seperti motivasi belajar, tingkat kehadiran, aktivitas organisasi, dan faktor psikologis guna meningkatkan akurasi model prediksi. Ketiga, penggunaan metode optimasi atau algoritma lain seperti Random Forest, Support Vector Machine, atau Deep Learning dapat dipertimbangkan untuk membandingkan dan meningkatkan performa model. Terakhir, perlu dilakukan pengujian pada dataset yang lebih besar dan lintas program studi agar model yang dihasilkan memiliki tingkat generalisasi yang lebih baik.

REFERENSI

- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing Undergraduate Students' Performance Using Educational Data Mining. *Computers & Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting Achievement and Providing Support Before STEM Majors Begin to Fail. *Computers & Education*, 158, 103999. <https://doi.org/10.1016/j.compedu.2020.103999>
- Burgos, C., Campanario, M. L., de la Peña, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data Mining for Modeling Students' Performance: A Tutoring Action Plan to Prevent Academic Dropout. *Computers and Electrical Engineering*, 66, 541–556. <https://doi.org/10.1016/j.compeleceng.2017.03.005>
- Chung, J. Y., & Lee, S. (2019). Dropout Early Warning Systems for High School Students Using Machine Learning. *Children and Youth Services Review*, 96, 346–353. <https://doi.org/10.1016/j.chilyouth.2018.11.030>
- Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the Effectiveness of Educational Data Mining Techniques for Early Prediction of Students' Academic Failure in Introductory Programming Courses. *Computers in Human Behavior*, 73, 247–256. <https://doi.org/10.1016/j.chb.2017.01.047>
- Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using Artificial Intelligence Methods to Assess Academic Achievement in Public High Schools of a European Union Country. *Heliyon*, 6(6), e04081. <https://doi.org/10.1016/j.heliyon.2020.e04081>
- Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational Data Mining: Predictive Analysis of Academic Performance of Public School Students in the Capital of Brazil. *Journal of Business Research*, 94, 335–343. <https://doi.org/10.1016/j.jbusres.2018.02.012>
- GIAN, F. (2014). *Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan Naive Bayes dan Algoritma C4.5 untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa*.
- Masci, C., Johnes, G., & Agasisti, T. (2018). Student and School Performance Across Countries: A Machine Learning Approach. *European Journal of Operational Research*, 269(3), 1072–1085. <https://doi.org/10.1016/j.ejor.2018.02.031>
- Rebai, S., Ben Yahia, F., & Essid, H. (2020). A Graphically Based Machine Learning Approach to Predict Secondary Schools Performance in Tunisia. *Socio-Economic Planning Sciences*, 70, 100724. <https://doi.org/10.1016/j.seps.2019.06.009>
- Rizvi, S., Rienties, B., & Ahmed, S. (2019). The Role of Demographics in Online Learning: A Decision Tree Based Approach. *Computers & Education*, 137, 32–47. <https://doi.org/10.1016/j.compedu.2019.04.001>
- Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting Academic Performance of Students from VLE Big Data Using Deep Learning

Models. *Computers in Human Behavior*, 104, 106189.
<https://doi.org/10.1016/j.chb.2019.106189>