

The Use Of Data Mining To Predict Student Study Period At Muhammadiyah University Sidoarjo By Using Naive Bayes Algorithm [Penggunaan Datamining Untuk Memprediksi Masa Studi Mahasiswa di Universitas Muhammadiyah Sidoarjo Dengan Algoritma Naive Bayes]

Muhammad Mursidil Arif¹⁾, Hamzah Setiawan ^{*,2)}

¹⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: 191080200034@umsida.ac.id, hamzah@umsida.ac.id

Abstract. *In the higher education, improving student performance and improving the quality of education is very important. The education system requires innovative ways to improve the quality of education, achieve the best results and minimize student failure rates. One of the innovative ways is to apply data mining to predict students' study period. The results of these predictions will help students or academic adviser to provide early warning or give more precise directions to each student, so that they can do the best things to increase the chances of graduating on time. In this study, 9 academic and non-academic variables were used, consisting of semester grade point index, Semesters 1, 2, 3 and 4, GPA, school origin (public/private), finance (constrained by financial problems or not), scholarship (whether get a scholarship or not), Student Affairs (active or not in the student program). The use of academic and non-academic data variables in this study aims to broaden the predictions of student graduation which are not only assessed from an academic point of view, but also look at non-academic factors. The data used is student's data for the 2017-2018 Informatics study program at the Muhammadiyah University of Sidoarjo. This data is obtained from the Directorate of Information Systems & Technology (DSTI) Muhammadiyah University of Sidoarjo as many as 200 data. Modelling using the naïve Bayes algorithm using Anaconda Navigator software with IDLE Jupyter Notebook and the Python programming language, after evaluation using the confusion matrix and accuracy score, the results obtained were 68% accuracy, precision value 0.67, recall 0.77 and f1-score 0.72. while the accuracy score evaluation value gets 67.35%.*

Keywords - *Datamining, Naïve Bayes, Confusion Matrix, Python, study period prediction*

Abstrak. *Dalam dunia pendidikan tinggi, peningkatan kinerja mahasiswa dan peningkatan kualitas Pendidikan merupakan salah satu hal yang penting. Sistem pendidikan membutuhkan cara-cara inovatif demi meningkatkan kualitas pendidikan dan mencapai hasil yang terbaik serta meminimalisir tingkat kegagalan mahasiswa. Salah satu cara inovatif itu adalah dengan menerapkan data mining untuk memprediksi masa studi mahasiswa. dari hasil prediksi tersebut nantinya akan membantu mahasiswa atau dosen wali untuk memberikan peringatan dini atau memberi arahan yang lebih tepat kepada setiap mahasiswa, sehingga dapat melakukan hal-hal yang terbaik untuk meningkatkan peluang lulus tepat waktu. Dataset yang digunakan merupakan data alumni program studi Informatika angkatan 2017-2018 Universitas Muhammadiyah Sidoarjo. Data ini didapatkan dari Direktorat Sistem & Teknologi Informasi (DSTI) Universitas Muhammadiyah Sidoarjo sebanyak 200 data. Permodelan menggunakan algoritma naïve bayes menggunakan software Anaconda Navigator dengan IDLE Jupyter Notebook dan Bahasa pemrograman python. Hasil yang didapatkan dari evaluasi menggunakan confusion matrix dan accuracy score didapatkan hasil accuracy 68%, nilai precision 0.67, recall 0.77 dan f1-score 0.72. sedangkan nilai evaluasi accuracy score mendapatkan 67.35%*

Kata Kunci - *Datamining, Naïve Bayes, Confusion Matrix, Python, study period prediction*

I. PENDAHULUAN

Dalam dunia pendidikan tinggi, peningkatan kinerja mahasiswa dan peningkatan kualitas pendidikan adalah salah satu hal yang penting. Hal tersebut juga berpengaruh terhadap penilaian kualitas akademik. Didalam peraturan BAN-PT (Organisasi dibawah kemedikti yang bertugas untuk akreditasi perguruan tinggi) pada pasal 2 tahun 2017 dijelaskan bahwa instrumen yang mempengaruhi akreditasi adalah capaian, dampak, mutu, dan produktifitas luaran yang bermanfaat langsung kepada masyarakat (Saputro dan Sari 2020).

Sistem pendidikan membutuhkan cara-cara inovatif demi meningkatkan kualitas pendidikan dan mencapai hasil yang terbaik serta meminimalisir tingkat kegagalan mahasiswa. dengan menggunakan Teknik data mining akan

didapatkan sebuah hasil prediksi yang dapat digunakan sebagai acuan untuk mendeteksi lebih dini tingkat kegagalan mahasiswa. sehingga tingkat hasil akademik akan naik dari tahun-ke tahun (Ashraf, Anwer, dan Gufran 2018).

Datamining merupakan salah satu metode yang sering digunakan untuk mendapatkan informasi penting bagi penggunanya [3]. Nama lain dari datamining adalah knowledge discovery in databases (KDD), yang artinya adalah proses pengumpulan data dan penggunaan data masa lampau, untuk mendapatkan informasi dari sebuah database [4]. Fungsi datamining secara umum adalah classification, clustering, Association, Regression, Forecasting, Sequence Analysis, dan Deviation Analysis [5]. cara kerja klasifikasi terbagi menjadi dua tahap penting, yaitu :

1. Pembuatan model yang nantinya disimpan sebagai prototype memori.
2. Menggunakan model yang disimpan untuk melakukan proses klasifikasi pada objek data lain dan menetapkan data tersebut sesuai kelasnya [6].

Penelitian yang dilakukan oleh sri widaningsih untuk menguji perbandingan metode datamining dengan membandingkan algoritma Naïve Bayes, C4.5, SVM dan KNN mendapatkan hasil bahwa algoritma yang memiliki tingkat keakuratan yang paling tinggi adalah naïve bayes dengan nilai akurasi 76.7%. Dalam penelitian tersebut, sri widaningsih menggunakan klasifikasi lulusan tepat waktu dan tidak tepat waktu sebagai variable target, dan menggunakan indeks prestasi semester (IPS) 3,4,5 dan 6 sebagai variable prediktor [7].

Dalam penelitian lainnya yang dilakukan oleh Mariana windarti tentang perbandingan 6 algoritma klasifikasi untuk prediksi masa studi mahasiswa, juga didapatkan hasil bahwa algoritma naïve bayes memiliki tingkat keakuratan yang paling tinggi dengan nilai akurasi 80.6%. Mariana windarti menggunakan variabel Index Prestasi (IP) semester 1,2 dan 3, penjurusan sekolah, jalur masuk perguruan tinggi sebagai predictor, dan variabel tahun lulus sebagai target [8].

Keunggulan algoritma naïve bayes dalam memprediksi masa studi mahasiswa bahkan membuat lila setyani menulis artikel tentang systematic review algoritma naïve bayes, yang mana didalam artikel tersebut membahas tentang penggunaan algoritma naïve bayes untuk memprediksi masa studi mahasiswa. didalam artikel tersebut juga dijelaskan bahwa keunggulan naïve bayes adalah : mudah difahami, pengkodean yang sederhana, dan kecepatan dalam perhitungan [5].

Dari penelitian terdahulu diatas, penggunaan algoritma naïve bayes untuk memprediksi masa studi memiliki tingkat akurasi yang lebih baik dibandingkan dengan algoritma yang lain, tetapi kebanyakan dari penelitian tersebut hanya terfokus pada banyaknya data, perbandingan algoritma, dan menggunakan variable dari data akademik mahasiswa. Atas dasar hal itu, pada penelitian ini menggunakan algoritma naïve bayes dan Bahasa pemrograman python dengan menggunakan variable akademik dan menambahkan variable non akademik.

Penelitian ini menggunakan 9 variable akademik dan non akademik, yang terdiri dari Indeks Prestasi Semester (IPS) 1,2,3 dan 4, IPK, asal sekolah (Negri/Swasta), Keuangan (Terkendala masalah keuangan atau tidak), Beasiswa (apakah mendapatkan beasiswa atau tidak), Kemahasiswaan (Aktif tidaknya mengikuti program kemahasiswaan). Penggunaan variable data akademik dan non akademik dalam penelitian ini bertujuan agar memperluas prediksi kelulusan mahasiswa yang tidak hanya dinilai dari segi akademik, tetapi juga melihat faktor non akademik. Dataset yang digunakan untuk penelitian ini merupakan data alumni program studi Informatika angkatan 2017-2018 Universitas Muhammadiyah Sidoarjo yang didapatkan dari Direktorat Sistem & Teknologi Informasi (DSTI) Universitas Muhammadiyah Sidoarjo sebanyak 200 data.

II. METODE

2.1 Data Mining

Data mining merupakan rangkaian proses yang dilakukan untuk memperoleh sebuah informasi dan pengetahuan dari database [6]. Nama lain dari datamining adalah knowledge discovery in databases (KDD), yang artinya adalah proses pengumpulan data dan penggunaan data masa lampau, untuk mendapatkan informasi dari sebuah database [4].

Adapun tahapan-tahapan yang dilakukan untuk datamining adalah data selection, preprocessing, datamining, interpretation dan evaluation [6].

2.2. Naïve Bayes

Naïve Bayes merupakan teknik probabilitas sederhana yang menerapkan teorema Bayes sebagai dasar dengan asumsi independen antar variabel. Ini memberikan probabilitas pada setiap objek untuk setiap kelas yang mungkin [9].

atau bisa disimpulkan bahwa naïve bayes menggunakan fitur independen sebagai model [6]

Dimana :

- C = Merupakan data yang kelasnya belum diketahui
- D = Hipotesa data dengan kelas spesifik
- $P(D|C)$ = Probabilitas hipotesa D berdasar kondisi C (Probabilitas posterior)
- $P(D)$ = Probabilitas hipotesa D (probabilitas prior)

$P(C|D)$ = probabilitas C berdasar pada kondisi hipotesa D

$P(C)$ = Probabilitas C

Dari penjelasan diatas, untuk menentukan kelas dapat dilakukan perbandingan nilai probabilitas dari sampel yang berada di kelas yang satu dengan nilai probabilitas dari sampel yang berada di kelas lain.

2.3. Tahapan Penelitian

Tahapan penelitian dilakukan dengan menjalankan langkah-langkah dari proses knowledge discovery in database (KDD), langkah-langkah tersebut yaitu : pengumpulan data, preprocessing data, permodelan naïve bayes, testing model, dan evaluasi[10]. lebih jelasnya digambarkan dengan diagram dibawah ini.

Gambar 1 Tahapan Penelitian

2.3.1 Pengumpulan Data

Pengumpulan data didapatkan dari Direktorat Sistem dan Teknologi Informasi (DSTI) Universitas Muhammadiyah Sidoarjo dengan menggunakan data alumni dari mahasiswa program studi informatika angkatan 2017-2018 sebanyak 200 data dengan 9 variabel prediktor 1 variabel target. Variabel prediktor terdiri dari Index Prestasi Semester (IPS) 1,2,3 dan 4, Index Prestasi Kumulatif (IPK), Asal sekolah (Negri/Swasta), Keuangan (selalu lunas tepat waktu/tidak tepat waktu), Beasiswa (Mahasiswa mendapatkan beasiswa/tidak mendapatkan), Kemahasiswaan (Mahasiswa pernah mengikuti kegiatan yang diadakan oleh kemahasiswaan/tidak pernah mengikuti). Sedangkan variabel target adalah Kelulusan (tepat waktu/tidak tepat waktu).

2.3.2 Preprocessing Data

Preprocessing merupakan langkah yang dilakukan sebelum pembuatan model untuk mempermudah dan memperlancar pembuatan suatu model. Dari data yang sudah didapatkan dilakukan pembersihan untuk memastikan tidak ada data ganda, dan data yang hilang atau memiliki nilai kosong (null) setelah proses data selection [7].

Untuk mempermudah proses klasifikasi menggunakan naïve bayes, maka data yang sebelumnya berbentuk text dirubah ke bentuk numerik dengan acuan sebagai berikut :

- Asal sekolah : Negri (0), Swasta (1)
- Keuangan : Lunas (0), Tidak (1)
- Beasiswa : Dapat (0), Tidak (1)
- Kemahasiswaan : Mengikuti (0), Tidak (1)
- Kelulusan : Tepat Waktu (0), Tidak (1).

2.3.3 Permodelan Naïve Bayes

Dalam proses permodelan data mining dilakukan klasifikasi dengan menerapkan algoritma naïve bayes. cara kerja algoritma naïve bayes menggunakan teorema bayes sebagai dasar yang akan menghitung probabilitas setiap kelas dengan syarat suatu kelas tidak tergantung dengan kelas yang lain [4].

Cara untuk menentukan kelas dapat dilakukan perbandingan nilai probabilitas dari sampel yang berada di kelas yang satu dengan nilai probabilitas dari sampel yang berada di kelas lain. Untuk melakukan proses data mining, dalam penelitian ini menggunakan aplikasi Anaconda Navigator dengan tools Jupyter Notebook. Sedangkan untuk pemrogramannya menggunakan bahasa Python.

2.3.4 Testing Model

Tahapan selanjutnya yaitu testing model dan Evaluasi. Testing model dilakukan dengan menguji model yang sudah dibuat dengan 70% data training (X-Predict) dengan data testing sebanyak 30% (Y-Test). Dari hasil testing model tersebut, nantinya akan dilakukan proses evaluasi untuk mengetahui nilai dari model.

2.3.5 Evaluasi

Untuk mengetahui tingkat akurasi dari pengujian yang telah dilakukan, maka langkah selanjutnya yaitu evaluasi. pada pengujian ini evaluasi menggunakan metode Confusion matrix sebagaimana skema pada gambar dibawah ini :

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

Gambar 2 Skema Confusion Matrix

True Positive = Banyaknya data dengan nilai positif yang diprediksi model juga positif.

True Negative = Banyaknya data dengan nilai negatif, yang diprediksi model juga negatif.

False Positive = Banyaknya data dengan nilai negatif, tetapi diprediksi model positif.

False Negative = Banyaknya data dengan nilai positif, tetapi diprediksi model negatif

Selanjutnya, cara mengetahui tingkat akurasi dari model yang telah didapatkan maka dilakukan dengan menghitung TRUE POSITIVE (TP) + TRUE NEGATIVE (TN) dibagi semua nilai.

Sedangkan untuk mengetahui tingkat errornya maka yang dihitung adalah nilai FALSE POSITIVE (FP) + FALSE NEGATIVE (FN) dibagi semua nilai.

Semua proses evaluasi dilakukan dengan memanggil fungsi dari Bahasa pemrograman python dengan menggunakan software jupyter notebook [11].

2.4 Python

Bahasa pemrograman python menjadi salah satu Bahasa yang paling banyak digunakan untuk mengembangkan suatu program termasuk untuk tugas analisis data [10]. Python mudah dipelajari dan juga merupakan bahasa pemrograman yang populer, hal tersebut adalah hasil survei dari pierre carbonelle (2016) yang dikutip oleh Mochammad kausar didalam artikelnya menunjukkan bahwa Bahasa python adalah Bahasa paling populer ke 2 setelah java [12].

3 III. Hasil dan Pembahasan

3.1 Pengumpulan Data

Dari analisis data yang sudah dilakukan pada bab sebelumnya, langkah selanjutnya adalah menjelaskan hasil dan pembahasan dari penelitian yang berjudul Penggunaan datamining untuk memprediksi masa studi mahasiswa universitas Muhammadiyah sidoarjo dengan algoritma naïve bayes. Proses pengumpulan data pada penelitian ini dilakukan dengan menggunakan data alumni program studi informatika universitas Muhammadiyah sidoarjo angkatan 2017-2018 yang didapatkan dari Direktorat Sistem dan Teknologi Informasi (DSTI) Universitas Muhammadiyah Sidoarjo sebanyak 200 data dengan 9 atribut numerik dan 1 label output. Pada gambar 3 berikut adalah tampilan dataset yang akan digunakan.

	ips1	ips2	ips3	ips4	ipk	sekolah asal	keuangan	beasiswa	kemahasiswaan	kelulusan
0	3.38	3.41	3.81	3.71	3.5775	1	0	1	1	1
1	3.42	3.29	3.58	3.47	3.4400	1	0	1	1	0
2	3.32	3.30	3.31	3.36	3.3225	0	0	1	1	1
3	3.32	3.26	3.17	3.25	3.2500	0	0	1	1	1
4	3.65	3.38	3.64	3.64	3.5775	1	0	1	1	1
5	3.43	3.22	3.60	3.35	3.4000	1	0	1	1	0
6	3.37	3.30	3.57	3.34	3.3950	1	0	1	1	1
7	3.47	3.28	3.47	3.23	3.3625	1	0	1	1	1
8	3.47	3.51	3.68	3.49	3.5375	1	0	1	0	0
9	3.37	3.35	3.60	3.32	3.4100	1	0	1	1	1

Gambar 3 Tampilan dataset

Dari 10 field diatas, 9 field pertama yaitu ips1, ips2, ips3, ips4 (Indeks Prestasi Semester 1-4), IPK, asal sekolah, keuangan, beasiswa, dan kemahasiswaan, akan dijadikan sebagai variable prediktor. Sedangkan field kelulusan akan menjadi variable target.

3.2 Proses Permodelan Naïve Bayes dengan Python

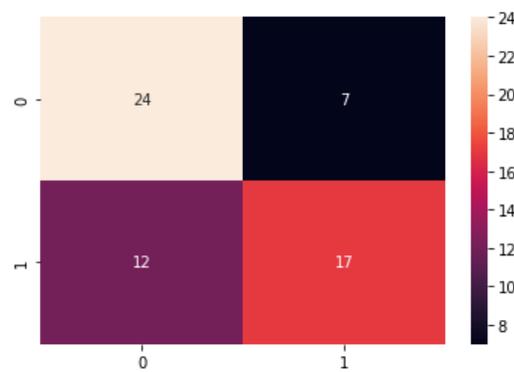
Proses klasifikasi naïve bayes dilakukan menggunakan aplikasi anaconda dengan tools jupyter notebook dan python sebagai bahasa pemrogramannya. yang pertama kali harus dilakukan adalah menyiapkan data dari Microsoft excel dengan format .xlsx, setelah itu mengimport library python yang diperlukan seperti (*pandas, numpy, matplotlib*). Berikut adalah langkah-langkah klasifikasi naïve bayes dengan python pada tools jupyter notebook :

- Dari dataset yang sebelumnya sudah dipersiapkan, diimport ke dalam tools jupyter notebook
- Langkah selanjutnya adalah membaca dataset dengan format .xlsx menggunakan library pandas
- Setelah file berhasil dibaca, selanjutnya adalah memastikan tidak ada data yang bernilai kosong (NULL), dan memastikan semua data bertipe numerik untuk mempermudah pembuatan model naïve bayes.
- Selanjutnya adalah pembagian antara data yang digunakan untuk training model dan data yang digunakan untuk testing model. dalam penelitian ini dilakukan pembagian dengan perbandingan 7:3 (70% data training – 30% data testing). Dari 200 dataset yang digunakan, didapatkan 140 data yang akan dijadikan data training, dan 60 data untuk dijadikan data testing

- e. Setelah pembagian data testing dan data training, dataset dipisahkan antara variable prediktor (x) dan variable target (y)
- f. Setelah semua data sudah siap, langkah berikutnya adalah membuat model dengan mengimport library gaussian naïve bayes dan memasukkan data training untuk mendapatkan model naïve bayes.
- g. Model yang didapatkan dari proses sebelumnya akan dilanjutkan dengan testing model menggunakan data testing yang sudah disiapkan sebelumnya.
- h. Langkah terakhir adalah evaluasi hasil dari proses testing model. Proses evaluasi model menggunakan confusion matrix untuk menghitung tingkat akurasi dari model naïve bayes.

3.2 Hasil Pengujian Confusion Matrix

Setelah dilakukan proses permodelan naïve bayes dengan bahasa pemrograman python pada tools jupyter notebook, maka didapatkan hasil sebagai berikut :



Gambar 4 Confusion Matrix

Hasil evaluasi menggunakan confusion matrix digambarkan pada gambar 4. Dimana dari 60 data testing, pada class 0 (Lulus tepat waktu) terdapat 24 data yang masuk kelas True Positive (TP) dan 7 data error atau masuk kelas True Negatif (TN). Sedangkan class 1 (Tidak tepat waktu) terdapat 12 data yang masuk kelas False Positive (FP), dan 17 data yang masuk kelas False Negatif (FN).

	precision	recall	f1-score	support
Tepat	0.67	0.77	0.72	31
Tidak Tepat	0.71	0.59	0.64	29
accuracy			0.68	60
macro avg	0.69	0.68	0.68	60
weighted avg	0.69	0.68	0.68	60

Gambar 5 Classification Report

Setelah mendapatkan nilai dari confusion matrix, pengujian model naïve bayes pada penelitian ini mendapatkan hasil accuracy 68%, dengan precision 0.67, recall 0.77 dan F1-score 0.72. Sedangkan accuracy score mendapatkan nilai 68.33%.

4 IV. SIMPULAN

Pada penelitian penggunaan datamining untuk memprediksi masa studi mahasiswa universitas Muhammadiyah sidoarjo dengan algoritma naïve bayes ini, didapatkan model dengan nilai accuracy score sebesar 68.33%. Sedangkan evaluasi menggunakan confusion matrix didapatkan hasil accuracy 68% dengan precision 0.67, recall 0.77 dan f1-score 0.72 untuk kelas tepat waktu, dan precision 0.71, recall 0.59, f1-score 0.64 untuk kelas tidak tepat waktu. Penelitian ini menggunakan data dari alumni program studi informatika angkatan 2017-2018 sebanyak 200 data dengan 9 variabel prediktor dan 1 variabel target.

5 UCAPAN TERIMA KASIH

Ucapan terima kasih kepada orang tua saya yang telah mendoakan dan menyemangati saya untuk menyelesaikan penelitian ini dengan cepat. Terima kasih kepada Bapak dan Ibu Dosen Informatika Universitas Muhammadiyah Sidoarjo yang telah membimbing dan memberikan ilmu selama kuliah. Tidak lupa saya ucapkan terima kasih kepada teman-teman A1, yang telah membantu dan memberikan dukungan selama penelitian saya berlangsung. Terima kasih juga kepada adinda Fani Miftakhul Faradisa yang selalu mensupport dan memberikan masukan dalam pembuatan skripsi. Dan terakhir saya ucapkan terima kasih kepada kepala DSTI bapak Moh. Alfian Rosyid, S.Kom., M.Kom yang membantu memberikan data untuk kelancaran skripsi ini.

6 REFERENSI

- [1] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.
- [2] A. Ashraf, S. Anwer, and M. Gufran, "A Comparative Study of Predicting Student 's Performance by use of Data Mining Techniques," *Am. Sci. Res. J. Eng. Technol. Sci.*, vol. 44, no. 1, pp. 122–136, 2018.
- [3] a Saputro and R. Helilintar, "Perancangan Prediksi Prestasi Nilai Akademik Mahasiswa Menggunakan Metode K-Means Clustering," *Prosiding Semnas Inotek ... 2020*, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/172>.
- [4] E. Sutoyo and A. Almaarif, "Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 95–101, 2020.
- [5] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, p. 35, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [6] M. W. Amelia, A. S. M. Lumenta, and A. Jacobus, "Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes," *E-Journal Tek. Inform.*, vol. 11, no. 1, 2017.
- [7] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [8] M. Windarti and A. Suranadi, "Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining Untuk Prediksi Masa Studi Mahasiswa," *J. Telemat.*, vol. 1, no. 1, pp. 14–30, 2019.
- [9] H. A. Mengash, "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020, doi: 10.1109/ACCESS.2020.2981905.
- [10] M. R. Qisthiano, T. B. Kurniawan, E. S. Negara, and M. Akbar, "Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 987, 2021, doi: 10.30865/mib.v5i3.3030.
- [11] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216–231, 2019, doi: 10.1016/j.patcog.2019.02.023.
- [12] M. K. Sophan, A. Kurniawati, T. Informatika, and U. T. Madura, "Perancangan Aplikasi Learning By Doing Interaktif Untuk Design of Interactive Applications " Learning By Doing " To Support," vol. 5, no. 2, pp. 163–170, 2018, doi: 10.25126/jtiik.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.