

Rancang Bangun Aplikasi Untuk Memprediksi Prestasi Siswa Menggunakan Algoritma Naive Bayes

Rizky Harianda Efendi^{1*}, Denny Kurniadi², Elfi Tasrif², Khairi Budayawan²

¹Pendidikan Teknik Informatika, Fakultas Teknik, Universitas Negeri Padang

²Departemen Teknik Elektronika, Universitas Negeri Padang

*Corresponding Author: rizkyharianda02@gmail.com

ABSTRACT

Student achievement is a crucial parameter in evaluating the effectiveness of the education system. In an effort to predict student performance, there are many factors that need to be considered. The selection of the appropriate prediction method can provide valuable insights for schools or educational institutions to improve student performance. The method used in this research is the Naive Bayes algorithm, a classification approach that utilizes Bayes' theorem to make predictions. The data used involves the distance of students' residences and grades in categorized subjects. The Naive Bayes algorithm allows us to model the probability of each achievement label based on the available data. This research reveals the predicted outcomes of student achievement based on data from 12 students, obtained from the school management. The data has also been tested using RapidMiner, yielding an accuracy rate of 83.33%, with a very good class prediction precision of 100%, a good prediction precision of 75.00%, and a prediction requiring guidance precision of 100%.

Keywords: Student Achievement Prediction, Naive Bayes Algorithm, Classification, Data Accuracy

ABSTRAK

Prestasi siswa adalah parameter krusial dalam mengevaluasi efektivitas sistem pendidikan. Dalam usaha untuk memprediksi prestasi siswa, terdapat banyak faktor yang perlu dipertimbangkan. Pemilihan metode prediksi yang tepat dapat memberikan wawasan berharga bagi pihak sekolah atau lembaga pendidikan guna meningkatkan kinerja siswa. Metode yang digunakan dalam penelitian ini adalah algoritma Naive Bayes, sebuah pendekatan klasifikasi yang memanfaatkan teorema Bayes untuk membuat prediksi. Data yang digunakan melibatkan jarak tempat tinggal siswa dan nilai mata pelajaran yang telah dikelompokkan. Algoritma Naive Bayes memungkinkan kita untuk memodelkan probabilitas setiap label prestasi berdasarkan data yang tersedia. Penelitian ini mengungkapkan hasil prediksi prestasi dari 12 data siswa, berdasarkan informasi yang diperoleh dari manajemen sekolah. Data tersebut juga telah diuji menggunakan RapidMiner, menghasilkan nilai akurasi sebesar 83,33%, dengan precision kelas prediksi yang sangat baik sebesar 100%, prediksi baik sebesar 75,00%, dan prediksi yang memerlukan bimbingan sebesar 100%.

Kata kunci: Prediksi Prestasi Siswa, Algoritma Naive Bayes, Klasifikasi, Akurasi Data

For all articles published in ELEKTIF, © copyright is retained by the authors.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



PENDAHULUAN

Sekolah Menengah Kejuruan (SMK) adalah salah satu jenis lembaga pendidikan yang bertujuan menciptakan sumber daya manusia yang kompeten dalam bidangnya [1]. SMK berkomitmen memberikan siswa keterampilan dan keahlian yang dapat digunakan untuk bekerja atau melanjutkan pendidikan ke jenjang yang lebih tinggi. SMK 3 Padangsidempuan, sebagai salah satu SMK di daerah ini, berusaha meluluskan siswa-siswanya dengan prestasi akademik yang baik agar mereka memiliki kesempatan lebih baik untuk melanjutkan ke perguruan tinggi. Setiap tahun, SMK 3 Padangsidempuan berhasil meluluskan lebih dari 200 siswa.

Untuk memastikan bahwa siswa-siswa ini memiliki peluang yang baik dalam proses penerimaan perguruan tinggi, SMK 3 Padangsidempuan mengumpulkan data dari setiap siswa mulai dari seluruh siswa kelas 1 hingga kelas 3. Dalam konteks pendaftaran ke perguruan tinggi di mana nilai atau prestasi menjadi tolak ukur, beberapa metode dapat diterapkan untuk meramalkan prestasi siswa di SMK 3 Padangsidempuan. Salah satu pendekatan yang digunakan adalah metode klasifikasi, di mana penggunaannya bertujuan untuk menghasilkan informasi yang dapat membantu sekolah dalam mengarahkan, membimbing, dan memberikan motivasi kepada siswa yang belum memenuhi sejumlah faktor yang dibutuhkan untuk diterima di perguruan tinggi. Hal ini dilakukan dengan tujuan agar siswa memiliki kemampuan bersaing yang sebanding dengan calon mahasiswa lainnya.

Namun, dalam penelitian ini, ditemukan kendala karena belum ada pola yang dapat dipergunakan untuk meramalkan siswa yang belum memenuhi sejumlah faktor penerimaan di perguruan tinggi, dengan dasar menggunakan nilai mata pelajaran dan perilaku sehari-hari. Dari permasalahan tersebut, diperlukan solusi untuk menemukan pola yang dapat digunakan dalam memprediksi kemampuan siswa SMK 3 Padangsidempuan, baik dalam bersaing di dunia kerja maupun untuk masuk perguruan tinggi. Salah satu teknik pengolahan data yang dapat diaplikasikan adalah Data Mining.

Dalam proses pengolahan Data Mining ini, peneliti memanfaatkan metode klasifikasi dengan menggunakan algoritma Naïve Bayes. Pendekatan ini bertujuan untuk mengidentifikasi serta mengukur tingkat akurasi yang dihasilkan oleh algoritma tersebut [2]–[5]. Data Mining merujuk pada konsep yang digunakan untuk mengeksplorasi pengetahuan dalam basis data [6]. Proses Data Mining sendiri merupakan analisis data dari sudut pandang yang berbeda, dan menghasilkannya dalam bentuk informasi yang krusial untuk meningkatkan keuntungan, mengurangi biaya pengeluaran, atau bahkan keduanya [7], [8].

Data mining melibatkan upaya pencarian tren atau pola yang diinginkan dalam basis data yang besar, bertujuan untuk mendukung pengambilan keputusan di masa mendatang. Pola-pola ini diidentifikasi oleh perangkat khusus yang dapat menyediakan analisis data yang bermanfaat dan berwawasan, yang kemudian dapat diteliti secara lebih rinci, mungkin dengan menggunakan alat bantu pengambilan keputusan tambahan [9]. Dalam penelitian sebelumnya, penerapan Algoritma Naïve Bayes pada prediksi kompetensi lulusan SMK Buddhi Tangerang terbukti memberikan hasil yang memuaskan, mencapai tingkat akurasi sebesar 98% [1]. Penelitian lain juga menunjukkan bahwa tujuan klasifikasi data mining adalah untuk meningkatkan keandalan hasil yang diperoleh dari data. Penggunaan Confusion Matrix sebagai metode evaluasi digunakan untuk mengukur nilai akurasi, recall, dan presisi dengan baik [1].

METODE

Dalam rangka penelitian ini, langkah-langkah penelitian yang akan dijalankan mencakup review literatur, pengumpulan data, pra-pemrosesan data, klasifikasi menggunakan *Naïve Bayes Classifier*, dan pembuatan kesimpulan penelitian. Penelitian ini bertujuan untuk mengkategorikan prestasi siswa dengan melibatkan serangkaian tahapan, termasuk (1) mengumpulkan data, (2) menetapkan kompetensi, (3) melakukan perhitungan, dan (4) menguji data menggunakan alat data mining, yaitu RapidMiner [10].

1. Pengumpulan data

Dalam tahap pengumpulan data latih, informasi yang digunakan melibatkan data siswa Jurusan Teknik Jaringan Komputer di SMK Negeri 3 Padangsidempuan pada periode 2020-2023, yang telah menyelesaikan pendidikan mereka di sekolah. Data ini diperoleh melalui manajemen sekolah. Data tersebut kemudian disubjekkan pada proses klasifikasi menggunakan algoritma *Naïve Bayes*. Hasil dari proses klasifikasi ini menghasilkan suatu model yang kemudian akan diimplementasikan pada data uji, yakni data siswa kelas X yang masuk pada tahun 2022-2023.

2. Penentuan Kompetensi

Pada label kompetensi dari prediksi ada 4 kategori, range dari nilai ini menyesuaikan dari format data siswa yang ada. kategori dari prediksi sebagai berikut :

Tabel 1. Range nilai kompetensi

Range	Kompetensi
0 - 79,33	perlu bimbingan
79,34 - 80,66	cukup
80,67 - 82,00	baik
82,01 - 100	sangat baik

3. Perhitungan

Langkah-langkah dalam perhitungan *Naïve Bayes* melibatkan (a) persiapan dataset sebagai bahan pelatihan, (b) perhitungan probabilitas untuk setiap kelas dan fitur yang digunakan, (c) perhitungan probabilitas akhir, dan (d) perbandingan probabilitas akhir [11]–[14].

a. Atribut

Terdapat total 8 atribut yang digunakan dalam penelitian ini. Enam atribut telah mencakup berbagai mata pelajaran dan telah dikelompokkan. Sementara itu, dua atribut tambahan mencakup jarak tempat tinggal dan nilai rata-rata dari kelompok mata pelajaran. Daftar atribut yang digunakan dalam penelitian ini dapat ditemukan pada tabel berikut:

Tabel 2. Atribut yang digunakan dalam penelitian

<i>Atribut</i>	<i>keterangan</i>
A1	Jarak tempat tinggal
A2	Nilai sains dasar
A3	Nilai keahlian komputer
A4	Nilai etika dan moral
A5	Nilai bidang sosial
A6	Nilai kecakapan komunikasi
A7	Nilai keahlian pendukung
A8	Rata – rata dari nilai

b. Menghitung Priors (Probabilitas kelas)

Berdasarkan kelas yang terdapat pada data jika kita ingin menghitung probabilitas bahwa data termasuk ke dalam kelas tertentu (misalnya kelas Baik) maka rumusnya sebagai berikut:

$$\begin{aligned} \text{Rumus priors} &= \frac{P(\text{Atribut}|\text{Kelas}) \times P(\text{Kelas})}{P(\text{atribut})} \\ &= \frac{\text{Jumlah Data training dengan kelas } y}{\text{total data training}} \end{aligned}$$

c. Hitung Likelihood (Probabilitas Atribut | Kelas)

Setelah melakukan perhitungan probabilitas kelas kemudian langkah selanjutnya adalah melakukan perhitungan Probabilitas Atribut. adapun persamaannya sebagai berikut :

$$\text{Rumus likelihood} = \frac{\text{Jumlah Data training dengan nilai } x \text{ dan kelas } y}{\text{total data training dengan kelas } y}$$

d . Hitung Poesterior (Probabilitas Kelas | Attibut)

Probabilitas posterior dalam konteks teorema Bayes adalah probabilitas kelas atau peristiwa tertentu setelah informasi baru atau bukti telah diperoleh. ini adalah estimasi probabilitas tentang kejadian tertentu setelah pengamatan data. adapun perasamaannya sebagai berikut :

$$\text{Poesterior} = (\text{Priors Kelas } y \times \text{Likelihood Atribut1 } y \times \text{Likelihood Atribut2 } y \times \text{Likelihood Atribut3 } y \times \text{Likelihood Atribut4 } y \times \text{Likelihood Atribut5 } y \times \text{Likelihood Atribut6 } y \times \text{Likelihood Atribut7 } y \times \text{Likelihood Atribut8 } y).$$

Dari semua persamaan diatas untuk mencari poesterior dari satu kelas (y) . untuk mencari hasil dari kelas lain langkah yang di perlukan sama contoh jika kelas (y) adalah baik maka jika nilai poesterior kelas baik lebih besar dari sangat baik , cukup dan perlu bimbingan maka hasil dari prediksinya adalah baik.

HASIL DAN PEMBAHASAN

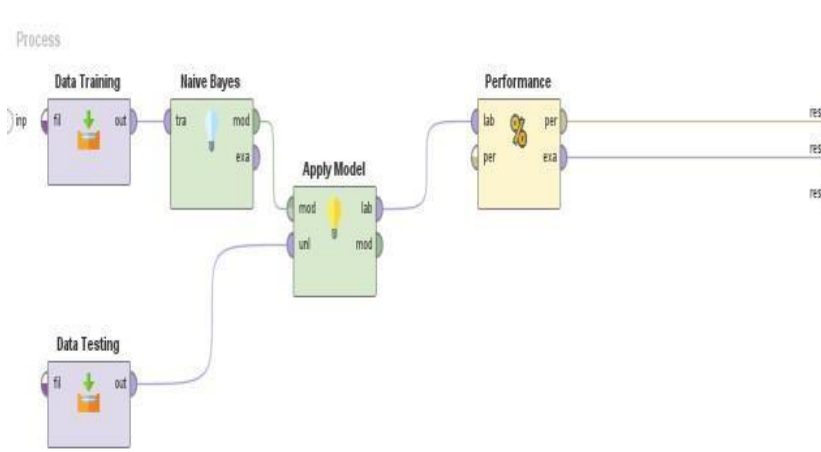
Hasil tabel di bawah ini di dapat dari konsep perhitungan pada algoritma *Naive Bayes*.dimana data yang diuji berjumlah 12 siswa , menggunakan 8 atribut yang menjadi tolak ukur untuk perhitungan dan 4 label kompetensi.

Tabel 3. Hasil perhitungan

No	A1	A2	A3	A4	A5	A6	A7	A8	Kompetensi
1	1-3km	81	80	82	83	81	81	81,33	Baik
2	1-3km	82	81	83	83	83	82	82,33	Sangat baik
3	3-5km	82	83	84	89	88	85	85,16	Sangat baik
4	3-5km	83	84	84	88	89	85	85,5	Sangat baik
5	1-3km	83	82	83	83	86	84	83,5	Sangat baik
6	1-3km	78	80	80	81	80	80	79,83	Baik
7	> 5km	83	85	84	90	87	85	85,66	Sangat baik
8	3-5km	82	83	82	83	88	81	83,16	Sangat baik
9	3-5km	80	84	79	83	81	79	81	Baik
10	3-5km	80	81	81	82	82	85	81,83	Baik
11	1-3km	79	80	79	82	80	84	80,66	Cukup
12	3-5km	58	55	52	61	51	67	57,33	Perlu bimbingan

Pengujian menggunakan aplikasi RapidMiner

Pengujian ini dilakukan dengan tujuan untuk mengevaluasi tingkat keakuratan klasifikasi kompetensi siswa menggunakan algoritma *Naive Bayes*. Data yang dianalisis merupakan data siswa kelas X Jurusan Teknik Komputer Jaringan SMK Negeri 3 Padangsidimpuan pada tahun ajaran 2022-2023. Data pelatihan terdiri dari data siswa yang telah lulus dengan jumlah total 108 siswa, sedangkan data pengujian melibatkan 12 siswa. Data siswa ini diperoleh dari manajemen sekolah dan mencakup 8 atribut yang menjadi fokus penelitian.



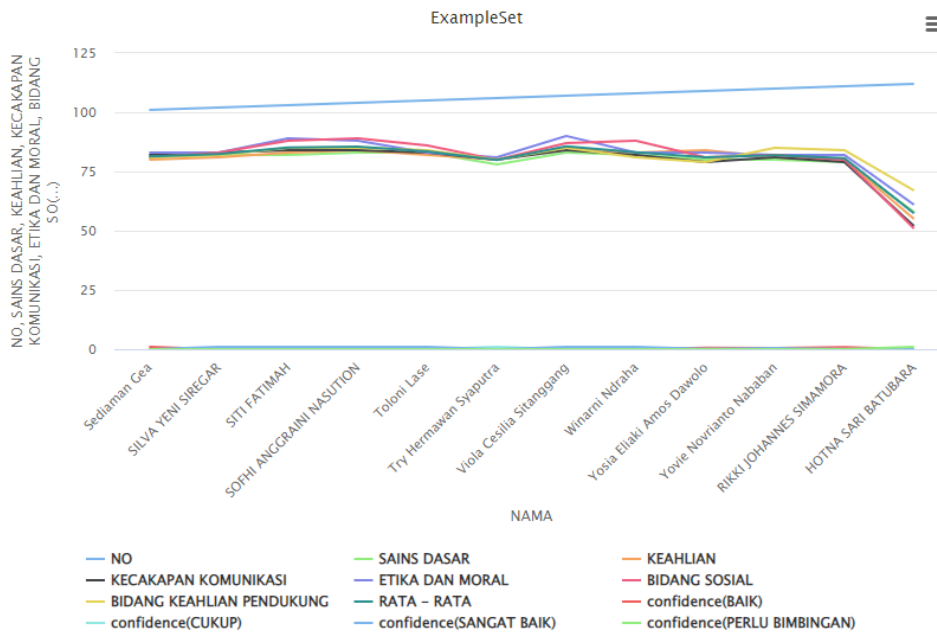
Gambar 1. Implementasi *Naive Bayes* untuk uji permormance

Gambar 1. merupakan pengaplikasian model dengan rapidminer untuk data siswa kelas X tahun 2022-2023.

	true BAIK	true SANGAT BAIK	true CUKUP	true PERLU BIMBING...	class precision
pred. BAIK	3	0	1	0	75.00%
pred. SANGAT BAIK	0	6	0	0	100.00%
pred. CUKUP	1	0	0	0	0.00%
pred. PERLU BIMBIN...	0	0	0	1	100.00%
class recall	75.00%	100.00%	0.00%	100.00%	

Gambar 2. Hasil pengujian pada rapidminer

Dengan menggunakan RapidMiner, hasil pengujian menunjukkan bahwa tingkat akurasi mencapai 83,33%, dengan rincian nilai presisi kelas sebagai berikut: Sangat Baik 100%, Baik 75,00%, Cukup 0,00%, dan Perlu Bimbingan 100%. Meskipun pada data pengujian terdapat satu kompetensi yang awalnya dianggap Cukup, namun setelah diuji menggunakan RapidMiner, kategori tersebut berubah menjadi Baik.



Gambar 3. Tampilan grafik pada rapidminer

Pada Gambar 3 cara melihat urutan nama siswa dari kiri merupakan contoh untuk hasil data testing nomor 1 itu berada pada sudut kiri siswa yang bernama Sediaman Gea.

PEMBAHASAN

Metode algoritma *Naive Bayes* digunakan dalam penelitian ini. Tujuan proyek ini adalah memberikan suatu wadah di mana prestasi siswa SMK Negeri 3 Padangsidempuan dapat diprediksi sejak dini. Apabila ada siswa yang hasil kompetensinya tidak memenuhi kriteria, pihak sekolah dapat melakukan bimbingan kepada siswa yang bersangkutan. Pada penelitian ini, digunakan 108 data siswa alumni SMK Negeri 3 Padangsidempuan tahun 2020-2022 sebagai data training, sedangkan untuk data testing menggunakan 12 data siswa kelas X tahun masuk 2022-2023.

Pada metode algoritma *Naive Bayes*, terdapat beberapa tahapan, mulai dari pengumpulan data, penentuan kompetensi, hingga perhitungan. Untuk penentuan range nilai kompetensi, menyesuaikan data dari pihak manajemen sekolah. Untuk tahap perhitungan, dimulai dari menghitung priors (probabilitas kelas), menghitung likelihood (probabilitas atribut|kelas), dan menghitung posterior (probabilitas kelas|atribut). Atribut yang digunakan dalam penelitian ini berjumlah 8, mulai dari jarak tempat tinggal, kelompok mata pelajaran sains dasar, kelompok mata pelajaran keahlian komputer, kelompok mata pelajaran etika dan moral, kelompok mata pelajaran bidang sosial, kelompok mata pelajaran kecakapan komunikasi, kelompok mata pelajaran bidang keahlian pendukung, hingga rata-rata nilai dari seluruh kelompok mata pelajaran.

Hasil perhitungan manual menggunakan algoritma *Naive Bayes* dari 12 data siswa sebagai data testing adalah 6 siswa kompetensi sangat baik, 4 siswa kompetensi baik, 1 siswa kompetensi cukup, dan 1 siswa kompetensi perlu bimbingan. Pada pengujian menggunakan tools rapidminer, dengan akurasi 83,00%, hasil yang diperoleh adalah 6 siswa kompetensi sangat baik, 5 siswa kompetensi baik, dan 1 siswa kompetensi cukup.

KESIMPULAN

Berdasarkan analisis dan pengujian data siswa SMK Negeri 3 Padangsidempuan Jurusan Teknik Komputer Jaringan tahun 2022-2023 dapat diambil kesimpulan sebagai berikut:

1. Penerapan metode data mining dengan menggunakan algoritma *Naive Bayes* telah berhasil dalam melakukan prediksi prestasi siswa di SMK 3.
2. Tingkat akurasi yang dihasilkan dari proses klasifikasi menggunakan algoritma *Naive Bayes* sangat dipengaruhi oleh data latih dan atribut yang digunakan sebagai acuan untuk menilai kompetensi siswa.
3. Dalam penelitian ini, terdapat 12 data siswa yang digunakan sebagai data uji, sementara data latih mencakup 108 siswa.
4. Hasil pengujian pada data siswa tahun 2022-2023 menunjukkan tingkat akurasi sebesar 83,00%.

DAFTAR PUSTAKA

- [1] L. Wydiastuty Kusuma, "Prediksi Kemampuan Lulusan SMK untuk Dapat Bersaing Di Dunia Kerja dengan Menggunakan Naive Bayes: Studi Kasus SMK Buddhi Tangerang," 2019. [Online]. Available: <https://jurnal.buddhidharma.ac.id/index.php/algor/index>
- [2] I. Algoritma *et al.*, "Implementasi Algoritma Naive Bayes Untuk Prediksi Persediaan Barang Rotan," *KOPERTIP : Scientific Journal of Informatics Management and Computer*, vol. 4, no. 1, pp. 28–34, Jun. 2020, doi: 10.32485/KOPERTIP.V4I1.112.

- [3] M. Asfi, N. Fitrianiingsih, D. Pembimbing, N. Skripsi, and C. Bayes, "Implementasi Algoritma *Naive Bayes* Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi," *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, no. 1, pp. 44–50, Sep. 2020, doi: 10.30743/INFOTEKJAR.V5I1.2536.
- [4] R. Y. Hayuningtyas, "Penerapan Algoritma *Naive Bayes* untuk Rekomendasi Pakaian Wanita," *Jurnal Informatika*, vol. 6, no. 1, pp. 18–22, Apr. 2019, doi: 10.31294/JI.V6I1.4685.
- [5] R. Aulia *et al.*, "Prediksi Perguruan Tinggi Negeri dengan Menggunakan Metode *Naive Bayes*," pp. 106–111, 2020.
- [6] M. Bakri, "Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K-Means," *Jurnal Teknoinfo*, vol. 11, no. 1, p. 6, 2017, doi: 10.33365/jti.v11i1.3.
- [7] E. Etriyanti, D. Syamsuar, D. Yesi, and N. Kunang, "Implementasi Data Mining Menggunakan Algoritma *Naive Bayes* Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa," *Telematika*, vol. 13, no. 1, pp. 56–67, Feb. 2020, doi: 10.35671/TELEMATIKA.V13I1.881.
- [8] D. Waru, R. W. Astuti, and N. Kahar, "Penerapan Data Mining Untuk Memprediksi Daya Serap Lulusan Siswa Menggunakan Algoritma *Naive Bayes*," *Arcitech: Journal of Computer Science and Artificial Intelligence*, vol. 1, no. 1, p. 57, Jun. 2021, doi: 10.29240/arcitech.v1i1.3294.
- [9] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naive Bayes* Classifier," *Creative Information Technology Journal*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.
- [10] A. Penelitian *et al.*, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 1, pp. 8–17, May 2021, doi: 10.25077/TEKNOSI.V7I1.2021.8-17.
- [11] J. A. Karlia, W. Nurmansyah, and M. Charitas, "Aplikasi Multi Platform Penerapan Algoritma *Naive Bayes* untuk Klasifikasi Keterlambatan Pembayaran Premi Asuransi," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 7, no. 1, pp. 9–15, Aug. 2022, doi: 10.30998/STRING.V7I1.11932.
- [12] J. N. Permana, R. Goejantoro, and S. Prangga, "Perbandingan Algoritma C4.5 Dan *Naive Bayes* Untuk Prediksi Ketepatan Waktu Studi Mahasiswa," *EKSPONENSIAL*, vol. 13, no. 2, pp. 161–170, Jan. 2023, doi: 10.30872/EKSPONENSIAL.V13I2.947.
- [13] N. Yustira, D. Witarsyah, and ..., "Implementasi Algoritma *Naive Bayes* Classification Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu," *eProceedings ...*, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16721%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16721/16429>
- [14] E. Fitriani, P. H. Susilo, and A. S. Budi, "Sistem Cerdas Prediksi Prestasi Belajar Menggunakan Algoritma *Naive Bayes* di MA Sains Roudlotul Qur'an Lamongan," *Generation Journal*, vol. 6, no. 1, pp. 58–67, Jan. 2022, doi: 10.29407/GJ.V6I1.16118.

Article submitted 2023-11-01. Resubmitted 2023-11-27. Final acceptance 2023-11-30. Final version published as submitted by the authors.