

Teknik Klasifikasi dalam Memprediksi Penerimaan Siswa Baru Menggunakan Metode Naïve Bayes

Classification Techniques in Predicting New Student Admission Using the Naïve Bayes Method

Suwayudhi¹, Eka Irawan², Bahrudi Efendi Damanik³
STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Article Info

Genesis Artikel:

Diterima, 23 September 2022

Direvisi, 13 Oktober 2022

Disetujui, 16 Oktober 2022

Kata Kunci:

Klasifikasi
Data Mining
Naïve Bayes
Prediksi
Siswa Baru

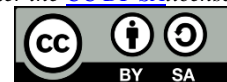
ABSTRAK

Penerimaan siswa baru merupakan proses pendaftaran kepada siswa baru masuk sekolah dan gerbang awal yang dilalui peserta didik didalam objek Pendidikan, kegiatan ini merupakan titik awal menentukan kelancaran tugas suatu sekolah, dengan dibantu Tenaga Pengajar serta dilengkapi sarana dan prasarana yang optimal dalam kegiatan belajar mengajar, menghasilkan siswa yang terampil dan berwawasan luas. Namun, ketidakpastian jumlah pendaftar turut mempengaruhi kebijakan yang kelak akan diambil. Oleh karena itu perlu dilakukan peramalan atau prediksi, untuk memperkirakan jumlah siswa yang kemungkinan akan mendaftar agar segala sesuatu nya dapat dipersiapkan oleh pihak sekolah. Pada penelitian ini proses prediksi penerima siswa baru akan menggunakan teknik klasifikasi dengan metode Naïve Bayes. Penelitian ini bertujuan untuk memprediksi naik turunnya jumlah siswa yang mendaftar dengan menggunakan metode Naïve Bayes. Data penelitian diperoleh dengan membagikan angket secara random kepada 200 responden (siswa) yang akan memasuki Sekolah Menengah Atas. Data di akumulasi dengan menggunakan bantuan Microsoft excel. Hasil yang diperoleh bahwa prediksi Tinggi class precision sebesar 100%, sedangkan prediksi Rendah class precision sebesar 94,23%. Kesimpulan nya bahwa pada kriteria Ekstrakurikuler, Biaya, dan Jarak perlu adanya perhatian dan perbaikan. Hal ini dikarenakan pada tidak minat dan prediksi rendah lebih tinggi dibandingkan minat dengan hasil prediksi tinggi.

ABSTRACT

Admission of new students is the registration process for new students entering school and the initial gate through which students enter the object of Education; this activity is the starting point for determining the smoothness of the tasks of a school, assisted by teaching staff and equipped with optimal facilities and infrastructure in teaching and learning activities, producing skilled and broad-minded students. However, the uncertainty of the number of registrants also influences the policies that will be taken in the future. Therefore, it is necessary to forecast or predict to estimate the number of students who are likely to register so that the school can prepare everything. In this study, the prediction process for new students will use a classification technique using the Naïve Bayes method. This study aims to predict the rise and fall of the number of students who register using the Naïve Bayes method. The research data was obtained by distributing questionnaires randomly to 200 respondents (students) who were about to enter high school. The data is accumulated using the help of Microsoft excel. The results obtained are that the prediction of high-class precision is 100%, while the prediction of low-class precision is 94.23%. The conclusion is that the extracurricular, cost and distance criteria need attention and improvement. This is because disinterest and low prediction are higher than interest with high prediction results.

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



Penulis Korespondensi:

Suwayudhi,
Program Studi Sistem Informasi
STIKOM Tunas Bangsa Pematangsiantar
Email: yudhi.11suwayudhi@gmail.com

1. PENDAHULUAN

Penerimaan siswa baru merupakan proses pendaftaran kepada siswa baru masuk sekolah dan gerbang awal yang dilalui peserta didik didalam objek Pendidikan, kegiatan ini merupakan titik awal menentukan kelancaran tugas suatu sekolah. Prediksi penerimaan siswa baru adalah memperkirakan banyaknya jumlah siswa terdaftar di tahun yang akan datang dengan perhitungan ilmiah [1].

Banyak cabang ilmu komputer yang mampu memecahkan masalah yang kompleks. Hal ini terbukti dari banyaknya penelitian-penelitian yang sudah dilakukan di bidang komputer untuk menyelesaikan masalah komputasi, seperti sistem pendukung keputusan [2]–[7], bidang jaringan syaraf tiruan [8]–[17], hingga di bidang data mining [18]–[24]. Data mining merupakan proses pengumpulan data dari beberapa informasi secara algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran, dan sistem manajemen suatu database [25]. Data mining klasifikasi dilakukan dengan pengklasifikasian data sesuai dengan target, variable, dan kategori dalam penyelesaian masalah.[26]. *Naïve Bayes* merupakan sebuah pengklasifikasian data secara *probabilistic* sederhana yang menghitung beberapa probabilitas dengan mengkombinasikan frekuensi nilai dari sebuah dataset. Proses metode ini dihitung dari kemiripan probabilitas kasus lama dengan kasus yang baru. [27].

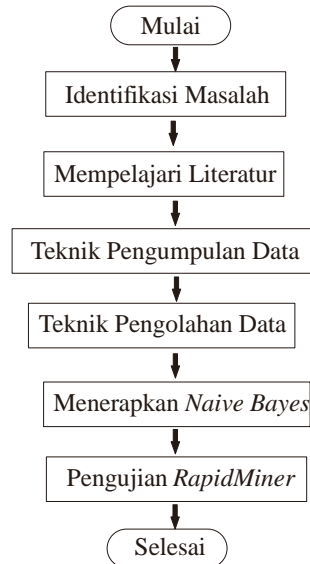
Pada tahun sebelumnya mengalami penurunan siswa baru yang diterima sehingga hal tersebut menjadi suatu permasalahan yang dihadapi oleh pihak sekolah. Berdasarkan kondisi tersebut, perlu adanya prediksi atau peramalan untuk mengetahui perolehan jumlah siswa baru di tahun yang akan datang. Dengan adanya prediksi pada jumlah siswa baru, pihak sekolah dapat mengetahui kebijakan yang diambil dalam meningkatkan jumlah siswa.

Berdasarkan uraian tersebut, maka dilakukan penelitian ini untuk mengklasifikasi penerimaan siswa baru pada Sekolah Menengah Atas menggunakan Algoritma *Naïve Bayes*.

2. METODE PENELITIAN

Rancangan penelitian ini dilakukan dengan pengamatan pada klasifikasi data siswa kedalam *Microsoft Excel*, kemudian data tersebut dilakukan suatu percobaan dengan menggunakan software Rapid Miner dengan metode *Naïve Bayes* untuk membuat model aturan data yang diperoleh dari Sekolah. Kemudian diharapkan dapat memprediksi penerimaan siswa baru pada sekolah menggunakan metode *Naïve Bayes*.

Bagan alir (*flowchart*) merupakan teknik analisis yang digunakan untuk menjelaskan aspek-aspek sistem informasi secara jelas, tepat dan logis [28]. Diagram alir model penelitian disajikan dalam rancangan pada gambar 1.



Gambar 1. Rancangan Penelitian

Adapun keterangan dari rancangan penelitian diatas adalah :

1. Identifikasi Masalah
Mengidentifikasi masalah yang menjadi penyebab naik atau menurunnya jumlah penerimaan siswa baru di Sekolah
2. Mempelajari Literatur
Pada penelitian ini, penulis harus didasari oleh rujukan yang digunakan untuk mendapatkan informasi terkait dan teori yang mendukung dalam penelitian.
3. Teknik Pengumpulan Data.
Teknik pengumpulan data merupakan teknik yang digunakan dalam mengumpulkan suatu informasi terkait yang dibutuhkan dalam melakukan penelitian. Adapun teknik pengumpulan data yang digunakan dalam penelitian ini yaitu:

- a. Studi dokumentasi
Sumber informasi yang diperoleh melalui studi dokumentasi, dengan mencari data yang berasal dari dokumen, buku, maupun jurnal.
 - b. Observasi
Observasi merupakan proses pencarian data yang dilakukan secara terbuka maupun tertutup pada pengambilan informasi pertama dengan mengamati objek yang diteliti [29].
 - c. Wawancara
Wawancara merupakan teknik pengambilan data dengan melakukan tanya jawab antara pengumpul data dengan narasumber [30].
4. Teknik Pengolahan Data
Penelitian ini melakukan pengolahan data dengan memilah data yang sesuai pada pokok permasalahan kemudian dimasukkan kedalam *Microsoft Excel 2010*.
 5. Menerapkan Metode
Naïve Bayes merupakan solusi yang digunakan pada penelitian ini dalam memprediksi jumlah penerimaan siswa baru di Sekolah
 6. Pengujian Aplikasi
Pada tahap ini, data yang telah di kelola sebelumnya akan di implementasi kedalam *Software Rapid Miner*.

3. HASIL DAN ANALISIS

3.1. Pengolahan Data

Adapun kriteria yang digunakan yaitu Fasilitas, Jurusan, Ekstrakurikuler Biaya, dan Jarak. Penulis memperoleh data dengan membagikan angket secara random sebanyak 200 responden kepada siswa. Data di akumulasi dengan menggunakan bantuan *software microsoft excel*. Data yang dikumpulkan merupakan data kuesioner yang dibagikan kepada para siswa yang akan memasuki sekolah menengah atas. Pengelompokan data sampel per parameter menjadi dua interval nilai seperti pada tabel berikut.

Tabel 1. Kelompok Nilai Kuesioner Per Parameter

No	Interval Nilai	Kelompok
1	1	Yes
2	0	No

Mengklasifikasi hasil menjadi dua interval nilai seperti pada tabel berikut :

Tabel 2. Klasifikasi Hasil Nilai Kuesioner

No	Interval Nilai	Klasifikasi
1	0-2	Rendah
2	3-5	Tinggi

Data awal diperoleh dengan *Microsoft Excel* sehingga menghasilkan data set pada tabel berikut :

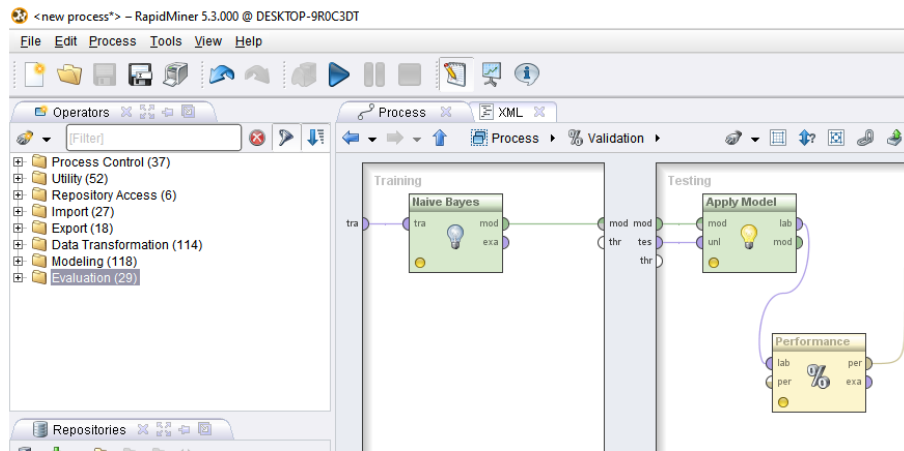
Tabel 3. Data Testing

Responden	Fasilitas	Jurusan	Ekstrakurikuler	Biaya	Jarak
R1	lengkap	Minat	Tidak Minat	Terjangkau	Terjangkau
R2	Tidak Lengkap	Tidak Minat	Minat	Terjangkau	Terjangkau
R3	lengkap	Minat	Minat	Tidak Terjangkau	Terjangkau
R4	lengkap	Minat	Tidak Minat	Terjangkau	Tidak Terjangkau
R5	lengkap	Tidak Minat	Minat	Tidak Terjangkau	Tidak Terjangkau
R6	Tidak Lengkap	Minat	Minat	Tidak Terjangkau	Tidak Terjangkau
R7	Tidak Lengkap	Tidak Minat	Tidak Minat	Terjangkau	Tidak Terjangkau
R8	lengkap	Minat	Minat	Terjangkau	Terjangkau
R9	lengkap	Minat	Minat	Terjangkau	Terjangkau
R10	lengkap	Minat	Minat	Tidak Terjangkau	Terjangkau
...
R198	lengkap	Minat	Minat	Terjangkau	Terjangkau
R199	lengkap	Minat	Minat	Terjangkau	Terjangkau
R200	lengkap	Minat	Minat	Terjangkau	Terjangkau

Tabel 3 merupakan data set yang telah di akumulasi menurut parameter. Setelah data tersebut di akumulasi, maka langkah selanjutnya mencari kriteria dan nilai probabilitas.

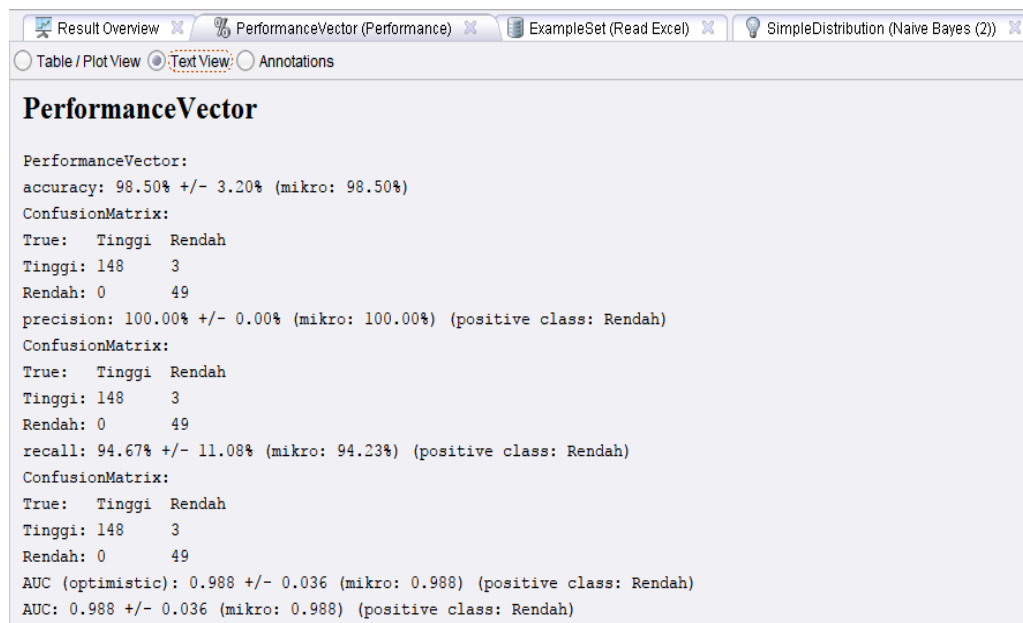
3.2. Hasil dan Pembahasan Aplikasi Rapid Miner Studio

Hasil dari pengujian data yang berjumlah 200 responden menggunakan software *Rapid miner* yaitu sebagai berikut: Proses *Validation* yang menghubungkan metode *naïve bayes* dengan apply model sehingga kita akan mengetahui nilai *Accuracy* nya. Gambar 2 berikut adalah *Process Validation* dari metode *naïve bayes*.



Gambar 2. *Process Validation, Naïve Bayes Method*

Setelah melalui proses diatas, maka hasil dari *performance vector* adalah *Accuracy* 98,50 %.



Gambar 3. *Performance Vector*

Dan nilai *Accuracy Performance* sebagai berikut :

accuracy: 98.50% +/- 3.20% (mikro: 98.50%)

	true Tinggi	true Rendah	class precision
pred. Tinggi	148	3	98.01%
pred. Rendah	0	49	100.00%
class recall	100.00%	94.23%	

Gambar 4. *Nilai Accuracy Performance*

Keterangan :

- a. Jumlah prediksi Tinggi dan kenyataannya Tinggi adalah 148 record
- b. Jumlah prediksi Rendah dan kenyataannya benar Tinggi adalah 0 record.
- c. Jumlah prediksi Tinggi dan kenyataannya benar Rendah adalah 3 record.
- d. Jumlah prediksi Rendah dan kenyataannya benar Rendah adalah 49 record

Pada gambar *Nilai Accuracy Performance* dijelaskan bahwa prediksi Tinggi memiliki jumlah 151 dengan *class precision* 100%, prediksi Rendah memiliki jumlah 49 dengan *class precision* 94,23 %. Sehingga total *Accuracy* yang diperoleh sebesar 98,50%.

4. KESIMPULAN

Berdasarkan hasil penyajian akhir maka dapat disimpulkan bahwa pada kriteria Ekstrakurikuler, Biaya, dan Jarak perlu adanya perhatian dan perbaikan. Dikarenakan pada tidak minat dan prediksi rendah lebih tinggi dibandingkan minat dengan hasil prediksi tinggi. Pada pengujian rapid miner dengan data kuesioner yang dibagikan pada 200 responden didapat bahwa prediksi Tinggi *class precision* 100%, prediksi Rendah *class precision* 94,23%. Pengujian dengan *rapid miner* merupakan perhitungan dari seluruh kriteria. Maka kesimpulan dari perhitungan tersebut bahwa mengalami kenaikan pada jumlah siswa baru yang akan masuk di Sekolah di tahun yang akan datang, dan perlu adanya perhatian terhadap kriteria yang kurang diminati oleh para siswa.

REFERENSI

- [1] A. Siti, "Analisis komparasi metode tsukamoto dan sugeno dalam prediksi jumlah siswa baru," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 2, pp. 57–63, 2016.
- [2] V. V. Sianipar, A. Wanto, and M. Safii, "Decision Support System for Determination of Village Fund Allocation Using AHP Method," *The IJICS (International Journal of Informatics and Computer Science) ISSN*, vol. 4, no. 1, pp. 20–28, 2020.
- [3] R. Simarmata, R. W. Sembiring, R. Dewi, A. Wanto, and E. Desiana, "Penentuan Masyarakat Penerima Bantuan Perbaikan Rumah di Kecamatan Siantar Barat Menggunakan Metode ELECTRE," *Journal of Computer System and Informatics (JoSYC)*, vol. 1, no. 2, pp. 68–75, 2020.
- [4] R. Watrionthos, W. A. Ritonga, A. Rengganis, A. Wanto, and M. Isa Indrawan, "Implementation of PROMETHEE-GAIA Method for Lecturer Performance Evaluation," *Journal of Physics: Conference Series*, vol. 1933, no. 1, p. 012067, 2021.
- [5] S. R. Ningsih, D. Hartama, A. Wanto, I. Parlina, and Solikhun, "Penerapan Sistem Pendukung Keputusan Pada Pemilihan Objek Wisata di Simalungun," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2019, pp. 731–735.
- [6] N. Nasution, G. W. Bhawika, A. Wanto, N. L. W. S. R. Ginantra, and T. Afriliansyah, "Smart City Recommendations Using the TOPSIS Method," *IOP Conference Series: Materials Science and Engineering*, vol. 846, no. 1, pp. 1–6, 2020.
- [7] R. A. Hutasoit, S. Solikhun, and A. Wanto, "Analisa Pemilihan Barista dengan Menggunakan Metode TOPSIS (Studi Kasus: Mo Coffee)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 2, no. 1, pp. 256–262, 2018.
- [8] I. M. Muhamad, S. A. Wardana, A. Wanto, and A. P. Windarto, "Algoritma Machine Learning untuk penentuan Model Prediksi Produksi Telur Ayam Petelur di Sumatera," *Journal of Informatics, Electrical and Electronics Engineering*, vol. 1, no. 4, pp. 126–134, 2022.
- [9] M. Mahendra, R. C. Telaumbanua, A. Wanto, and A. P. Windarto, "Akurasi Prediksi Ekspor Tanaman Obat , Aromatik dan Rempah-Rempah Menggunakan Machine Learning," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 6, pp. 207–215, 2022.
- [10] R. Puspadini, A. Wanto, and N. Arminarahmah, "Penerapan ML dengan Teknik Bayesian Regulation untuk Peramalan," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 3, pp. 147–155, 2022.
- [11] N. L. W. S. R. Ginantra, A. D. GS, S. Andini, and A. Wanto, "Pemanfaatan Algoritma Fletcher-Reeves untuk Penentuan Model Prediksi Harga Nilai Ekspor Menurut Golongan SITC," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 679–685, 2022.
- [12] N. Arminarahmah, S. D. Rizki, O. A. Putra, U. Islam, K. Muhammad, and A. Al, "Performance Analysis and Model Determination for Forecasting Aluminum Imports Using the Powell-Beale Algorithm," *IJISTECH (International Journal of Information System & Technology)*, vol. 5, no. 5, pp. 624–632, 2022.
- [13] N. L. W. S. R. Ginantra *et al.*, "Performance One-step secant Training Method for Forecasting Cases," *Journal of Physics: Conference Series*, vol. 1933, no. 1, pp. 1–8, 2021.
- [14] A. Perdana, S. Defit, and A. Wanto, "Optimalisasi Parameter dengan Cross Validation dan Neural Back-propagation

- Pada Model Prediksi Pertumbuhan Industri Mikro dan Kecil,” *Jurnal Sistem Informasi Bisnis*, vol. 01, no. 11, pp. 34–42, 2021.
- [15] N. L. W. S. R. Ginantra, M. A. Hanafiah, A. Wanto, R. Winanjaya, and H. Okprana, “Utilization of the Batch Training Method for Predicting Natural Disasters and Their Impacts,” *IOP Conf. Series: Materials Science and Engineering*, vol. 1071, no. 1, p. 012022, 2021.
- [16] A. Wanto, S. Defit, and A. P. Windarto, “Algoritma Fungsi Pelatihan pada Machine Learning berbasis ANN untuk Peramalan Fenomena Bencana,” *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 254–264, 2021.
- [17] V. V. Utari, A. Wanto, I. Gunawan, and Z. M. Nasution, “Prediksi Hasil Produksi Kelapa Sawit PTPN IV Bahjambi Menggunakan Algoritma Backpropagation,” *Journal of Computer System and Informatics (JoSYC)*, vol. 2, no. 3, pp. 271–279, 2021.
- [18] N. Arminarahmah, A. D. GS, G. W. Bhawika, M. P. Dewi, and A. Wanto, “Mapping the Spread of Covid-19 in Asia Using Data Mining X-Means Algorithms,” *IOP Conf. Series: Materials Science and Engineering*, vol. 1071, no. 1, p. 012018, 2021.
- [19] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, “COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm,” *Journal of Physics: Conference Series*, vol. 1783, no. 1, p. 012027, 2021.
- [20] N. A. Febriyati, A. D. GS, and A. Wanto, “GRDP Growth Rate Clustering in Surabaya City uses the K- Means Algorithm,” *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 276–283, 2020.
- [21] M. A. Hanafiah and A. Wanto, “Implementation of Data Mining Algorithms for Grouping Poverty Lines by District/City in North Sumatra,” *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 315–322, 2020.
- [22] T. H. Sinaga, A. Wanto, I. Gunawan, S. Sumarno, and Z. M. Nasution, “Implementation of Data Mining Using C4.5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM,” *Journal of Computer Networks, Architecture, and High-Performance Computing*, vol. 3, no. 1, pp. 9–20, 2021.
- [23] A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.
- [24] W. T. C. Gultom, A. Wanto, I. Gunawan, M. R. Lubis, and I. O. Kirana, “Application of The Levenberg Marquardt Method In Predict The Amount of Criminality in Pematangsiantar City,” *Journal of Computer Networks, Architecture, and High-Performance Computing*, vol. 3, no. 1, pp. 21–29, 2021.
- [25] R. Yanto and R. Khoiriah, “Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat,” *Creative Information Technology Journal*, vol. 2, no. 2, p. 102, 2015.
- [26] S. Faisal, “Klasifikasi Data Mining Menggunakan Algoritma C4.5 Terhadap Kepuasan Pelanggan Sewa Kamera Cikarang,” *Techno Xplore : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 4, no. 1, pp. 1–8, 2019.
- [27] S. M. Dewi, A. P. Windarto, and D. Hartama, “Penerapan Datamining Dengan Metode Klasifikasi Untuk Strategi Penjualan Produk Di Ud.Selamat Selular,” *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp. 617–621, 2019.
- [28] K. E. Setyaputri, Megawati, M. W. Fadholi, and F. C. Mukti, “Rancang Bangun Sistem Pelaporan a1 Berbasis Web Di Badan Pengawas Pemilu Kabupaten Brebes,” *Jurnal Teknik Informatika dan Sistem Informasi (JURTISI)*, vol. 1, no. 1, pp. 1–7, 2021.
- [29] A. Senika, R. Rasiban, and D. Iskandar, “Implementasi Metode Naïve Bayes Dalam Penilaian Kinerja Sales Marketing Pada PT. Pachira Distrinusa,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, p. 701, 2022.
- [30] A. G. P. Alistiani, Robby Rizky, Lili Sujai, “Implementasi Metode Naive Bayes Dan Klasifikasi Pegawai Terbaik Menggunakan Metode Naive Bayes,” *Situstika Fikunma*, vol. 8, no. 1, pp. 1–7, 2019.