

Penerapan Data Mining Untuk Klasifikasi Kelayakan Nasabah Dalam Pengajuan Kredit Menggunakan Metode *Naives Bayes* dan *Random Forest*

¹Muhammad Zakaria, ²Andri

^{1,2}Fakultas Ilmu Komputer, Universitas Binadarma Palembang, Indonesia

¹muhamadzakaria21@gmail.com; ²andri@binadarma.ac.id

Article Info

Article history:

Received, 2023-09-04

Revised, 2023-10-02

Accepted, 2023-11-20

Kata Kunci:

Naives Bayes, Random Forest, Data Mining

Keywords:

Naives Bayes, Random Forest, Data Mining

ABSTRAK

Peranan dari data mining sangat penting dalam sebuah manajemen bisnis yang saat ini semakin kompetitif. Data Mining dapat menggunakan aset penting pada perusahaan sebagai data bisnis yang memiliki jumlah besar, sehingga dari data tersebut dapat diolah menjadi sebuah informasi. Dalam data mining memiliki beberapa teknik yang dapat digunakan, salah satunya ialah klasifikasi. Klasifikasi itu sendiri memiliki beberapa algoritma, diantaranya algoritma *Naives Baye dan Random Forest*. Keuntungan dalam menggunakan algoritma *Naives Bayes dan random forest* adalah algoritma ini menggunakan data pelatihan (*Training Data*) yang kecil dalam menentukan estimasi pada parameter yang diperlukan pada sebuah proses pengklasifikasian. Penelitian ini menggunakan parameter pinjaman yang dihasilkan, dan parameter tersebut dapat dibuat satu penilaian terhadap status pinjaman yang ada pada Bank Palembang. Sehingga dapat menentukan klasifikasi calon nasabah yang layak atau tidak untuk menerima pinjaman. Untuk mengetahui kelayakan pinjaman tersebut maka, diperlukan adanya sebuah peramalan yang akurat dengan nilai akurasi yang baik. Untuk mengetahui nilai akurasi tersebut dapat menggunakan teknologi di bidang *Data Mining*. Hasil penelitian menunjukkan nilai akurasi 78.4% terhadap data nasabah yang digunakan pada algoritma naives bayes.

ABSTRACT

The role of data mining is very important in managing businesses which are currently increasingly competitive. Data Mining can utilize important company assets as large amounts of business data, so that the data can be processed into information. In data mining there are various techniques that can be used, one of which is classification. The classification itself has several algorithms, one of which is the Naives Bayes algorithm. The advantage of using the Naives Bayes algorithm is that this algorithm uses small training data to determine parameter estimates needed in a classification process. This research uses the resulting loan parameters, and these parameters can be used to assess the status of existing loans at Bank Palembang. So you can determine the classification of potential customers who are eligible or not to receive a loan. To determine the suitability of a loan requires an accurate forecast with good precision. To find out the accuracy value, you can use technology in the field of Data Mining. The research results show an accuracy value of 78.4% for the customer data used.

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



Penulis Korespondensi:

Andri,

Program Studi Informatika,

Universitas Bina Darma

Email: jojo@gmail.com

1. PENDAHULUAN

Dalam perkembangan teknologi masa kini, banyaknya data menjadi sebuah permasalahan sekaligus kesempatan bagi sebuah instansi. Data menjadi permasalahan apabila tidak dapat disimpan, dikelola, maupun diproses dengan baik. Data yang selalu bermunculan setiap waktu akan terus menumpuk dan bila tidak didokumentasikan dengan baik, maka data tersebut akan menjadi tidak berguna untuk perusahaan. Sedangkan data menjadi sebuah kesempatan apabila dapat disimpan, dikelola dan diproses menjadi lebih berarti untuk instansi tersebut. Dengan adanya data, maka dapat ditemukan sebuah trend maupun struktur yang nantinya dapat dipergunakan untuk mendapatkan informasi di masa mendatang [1].

Bank Palembang merupakan salah satu perusahaan yang merasakan perkembangan teknologi. Bank Palembang merupakan badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak. Pada bank Palembang memiliki tugas utama yaitu mengembangkan sejumlah set model serta teknik dengan tujuan untuk memungkinkan pihak mereka atau badan keuangan yang dimiliki dalam menentukan kelayakan simpan pinjam calon nasabah.

Masalah yang sering terjadi ialah, bagaimana pola perilaku nasabah atau karakteristik nasabah yang dimiliki sehingga yang terjadi dapat menentukan pola atau status dari calon nasabah. Dalam pelaksanaannya, analisis pinjaman di Bank Palembang melaksanakan pengambilan data calon nasabah pinjaman dan selanjutnya data tersebut di analisis untuk menentukan kelayakan. Dalam penentuan analisa kelayakan calon nasabah penerima pinjaman di Bank Palembang, bagian analis akan meminta data kepada Bank Indonesia melalui web serta analisis tersebut akan menganalisisnya secara mandiri. Analisis yang dilakukan terhadap data pinjaman bank digunakan dengan tujuan agar dapat meminimalisasi resiko nasabah yang memiliki penilaian kurang baik sehingga berdampak pada keterlambatan dalam membayar pinjaman.

Peranan dari data mining sangat penting dalam sebuah manajemen bisnis yang saat ini semakin kompetitif [2]. Data Mining dapat menggunakan aset penting pada perusahaan sebagai data bisnis yang memiliki jumlah besar, sehingga dari data tersebut dapat diolah menjadi sebuah informasi [3]. Dalam data mining memiliki beberapa teknik yang dapat digunakan, salah satunya ialah klasifikasi. Klasifikasi itu sendiri memiliki beberapa algoritma, diantaranya algoritma Naives Bayes [4]. Random forest merupakan peramalan pada kondisi yang dimiliki oleh suatu objek[5].

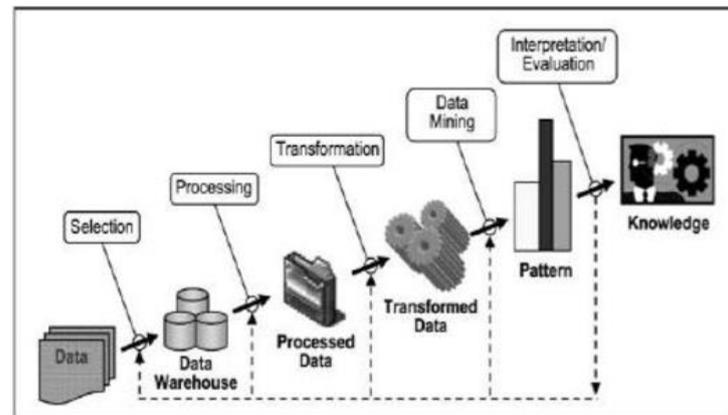
Keuntungan dalam menggunakan algoritma Naives Bayes adalah algoritma ini menggunakan data pelatihan (Training Data) yang kecil dalam menentukan estimasi pada parameter yang diperlukan pada sebuah proses pengklasifikasian [6].

Penelitian ini menggunakan parameter pinjaman yang dihasilkan, dan parameter tersebut dapat dibuat satu penilaian terhadap status pinjaman yang ada pada Bank Palembang. Sehingga dapat menentukan klasifikasi calon nasabah yang layak atau tidak untuk menerima pinjaman.

Untuk mengetahui kelayakan pinjaman tersebut maka, diperlukan adanya sebuah peramalan yang akurat dengan nilai akurasi yang baik. Untuk mengetahui nilai akurasi tersebut dapat menggunakan teknologi di bidang Data Mining [7]. Penelitian ini menentukan klasifikasi pemutusan kredit nasabah dan mengevaluasi kinerja metode Algoritma Naives Bayes dan Random Forest untuk melihat nilai support dan confident yang dihasilkan dari metode tersebut.

2. METODE PENELITIAN

Knowledge Discovery in Database (KDD) merupakan sebuah proses dalam menentukan sebuah informasi yang dapat berguna serta menghasilkan pola-pola yang terdapat dalam data [8]. Informasi tersebut terkandung pada basis data memiliki ukuran yang besar sehingga sebelumnya tidak dapat diketahui dan potensial bermanfaat. *Data mining* ialah sebuah cara atau tahapan langkah dari sekumpulan proses *iterative KDD* [9].



Sumber: Kursni, Emha Thaufiq Luthfi, 2009:7

Gambar 1. Tahapan Dalam KDD

Adapun tahapan pada proses *Knowledge Discovery in Database* (KDD) ialah sebagai berikut :

1. *Data Selection*

Proses seleksi data ini dilakukan untuk melakukan pemilihan himpunan data, serta menciptakan himpunan data target, atau memfokuskan pada subset *variable* (sampel data) dimana penemuan (*discovery*) akan dilakukan [10].

2. *Pre-Processing* dan *Cleaning Data*

Pemrosesan data dan pembersihan data dilakukan agar membuang data yang didalamnya terdapat data yang tidak konsisten atau *noise*, duplikat data atau data ganda, serta memperbaiki kesalahan data, hal tersebut bisa diperkaya dengan mengumpulkan data *eksternal* yang relevan [11].

3. *Transformation*

Proses ini melakukan transformasi atau penggabungan data ke dalam sebuah parameter yang elbih tepatnya dilaksanakan proses mining [12].

4. *Data Mining*

Data Mining ialah proses melaksanakan penarikan data sehingga menghasilkan sebuah pola maupun klasifikasi dalam sebuah asosiasi [13].

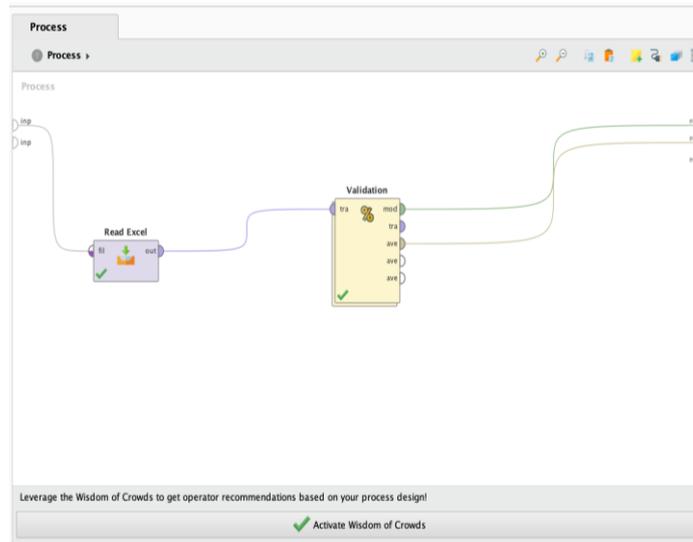
5. *Interpretation / Evaluasi*

Evaluasi menguji proses atau pola yang dihasilkan dengan melakukan analisis nilai akurasi yang ada [14].

3. HASIL DAN PEMBAHASAN

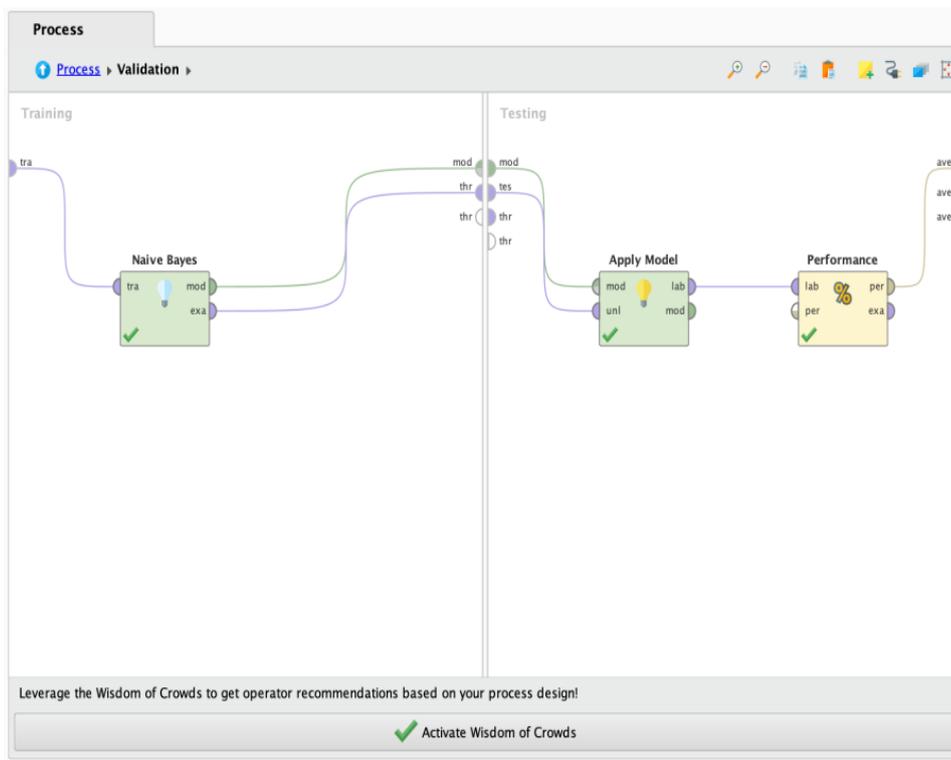
Penelitian ini melakukan proses klasifikasi pola nasabah yang akan melakukan pinjaman di Bank Palembang. Parameter yang digunakan akan berpengaruh pada nilai akurasi, adapun parameter yang digunakan ialah penghasilan, pekerjaan dan status nasabah di Bank Indonesia.

Pertama yang dilakukan ialah melakukan struktur model dengan menghubungkan data training pada excel pada bagian atribut (read excel) dan disambungkan dengan atribut validation. Hal tersebut dilakukan untuk mendapatkan nilai akurasi dari algoritma naives bayes. Desain model naives bayes menggunakan bantuan tools RapidMiner yang ditunjukkan pada gambar dibawah ini.



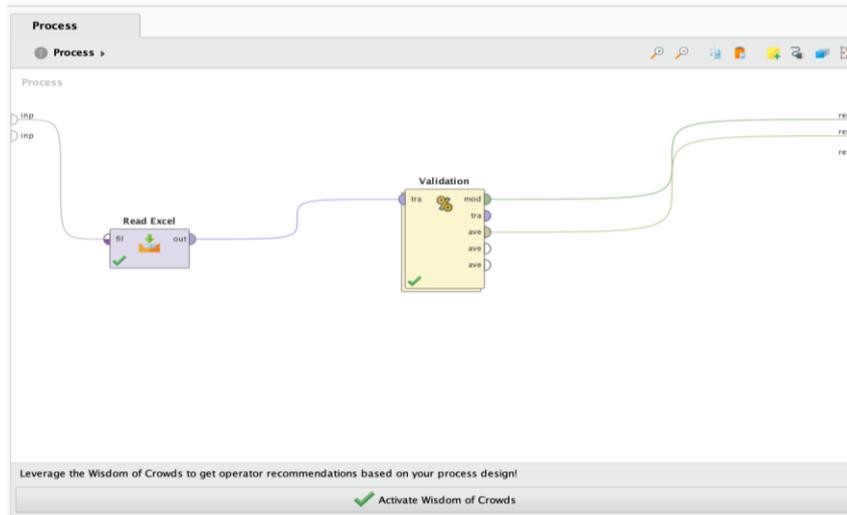
Gambar 2. Cross Validation

Selanjutnya operator naives bayes yang digunakan sebagai algoritma pada data mining *Classification*. Maka selanjutnya menggunakan pengabungan operator Apply Model dan operator Performance yang diimplementasikan untuk menghasilkan proses pengolahan data mining. Adapun model operator sub proses dapat di lihat pada Gambar sebagai berikut :



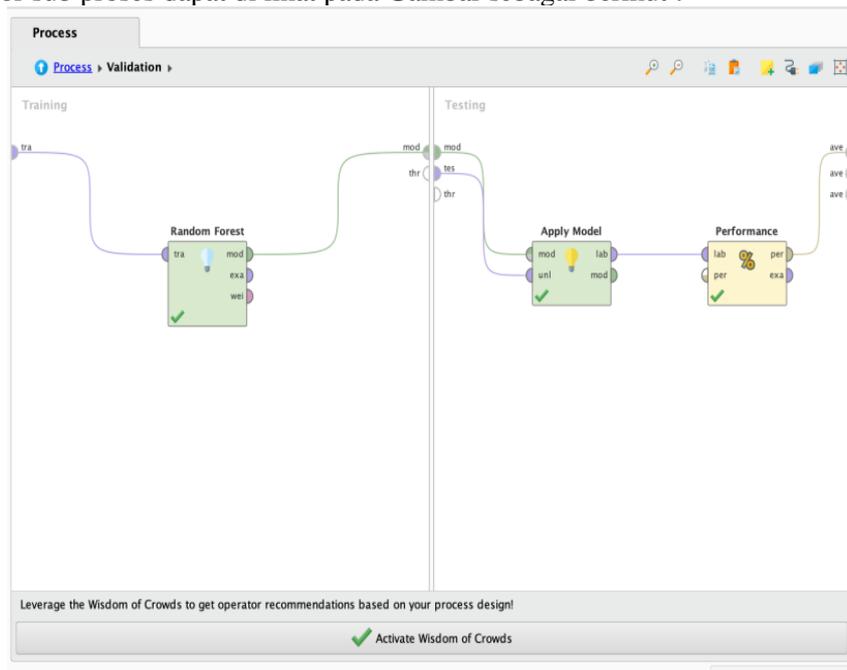
Gambar 3. Rapidminer Model Sub Proses Clasification Naives Bayes

Pertama yang dilakukan ialah melakukan struktur model dengan menghubungkan data training pada excel pada bagian atribut (read excel) dan disambungkan dengan atribut validation. Hal tersebut dilakukan untuk mendapatkan nilai akurasi dari algoritma random forest . Desain model random forest menggunakan bantuan tools RapidMiner yang ditunjukkan pada gambar dibawah ini. Desain modelnya yang terdapat pada tools RapidMiner ditunjukkan pada gambar dibawah ini :



Gambar 3. Cross Validation

Selanjutnya operator naives bayes yang digunakan sebagai algoritma pada data mining *Classification*. Maka selanjutnya menggunakan penggabungan operator Apply Model dan operator Performance yang diimplementasikan untuk menghasilkan proses pengolahan data mining. Adapun model operator sub proses dapat di lihat pada Gambar sebagai berikut :



Gambar 4. Rapidminer Model Sub Proses Clasifikasi Random Forest

Tujuan penelitian ini ialah salah agar dapat mengetahui nilai akurasi dari algoritma yang digunakan yaitu algoritma klasifikasi naives bayes dan random forest. Penelitian ini di lakukan untuk mengetahui klasifikasi pola nasabah yang akan melakukan pinjaman. Penggunaan parameter berpengaruh pada hasil akurasi dan model yang akan dihasilkan oleh algoritma naives bayes dan random forest.

Setelah dilaksanakannya proses distraksi maka diperoleh nilai akurasi dari algoritma naives bayes. Hasil yang diperoleh merupakan nilai akurasi, *recall*, dan *precision* algoritma naives bayes. Adapun hasil yang ditampilkan pada tools rapid miner ialah sebagai berikut :

The screenshot shows the 'Performance' tab in Rapidminer for a Naives Bayes model. The 'Criterion' is set to 'accuracy'. The overall accuracy is 78.10%. The confusion matrix table is as follows:

	true Outstanding	true Paid	class precision
pred. Outstanding	0	0	0.00%
pred. Paid	23	82	78.10%
class recall	0.00%	100.00%	

Gambar 5. Rapidminer Performance Naives Bayes

Dari gambar diatas dapat diketahui bahwa nilai akurasi yang ada ialah 78,10%, class recall untuk outstanding 0%, class recall paid ialah 100% serta precision 78,10%.

Setelah dilaksanakannya proses distraksi maka diperoleh nilai akurasi dari algoritma random forest. Hasil yang diperoleh merupakan nilai akurasi, *recall*, dan *precision* algoritma random forest. Adapun hasil yang ditampilkan pada tools rapid miner ialah sebagai berikut :

The screenshot shows the 'Performance' tab in Rapidminer for a Random Forest model. The 'Criterion' is set to 'accuracy'. The overall accuracy is 76.19%. The confusion matrix table is as follows:

	true Outstanding	true Paid	class precision
pred. Outstanding	0	2	0.00%
pred. Paid	23	80	77.67%
class recall	0.00%	97.56%	

Gambar 6. Rapidminer Performance Random Forest

Dari gambar diatas dapat diketahui bahwa nilai akurasi yang ada ialah 76,08%, class recall untuk outstanding 0%, class recall paid ialah 97,56% serta precision 77,67%.

Setelah diproses analisis dan diperoleh hasil pengujian data training dari algoritma naives bayes dan algoritma random forest, maka kedua algoritma tersebut memiliki hasil yang perbandingan sehingga dapat ditarik kesimpulan mana algoritma yang akurat. Adapun hasil dari kedua algoritma tersebut dapat dilihat sebagai berikut :

Tabel 1. Perbandingan Algoritma naives bayes dan Algoritma random forest

No.	Algoritma Naives Bayes	Algoritma Random Forest
1.	Memiliki nilai akurasi 27,4% untuk data testing.	Memiliki nilai akurasi 78,4% untuk data testing.
2.	Memiliki nilai akurasi 78.10% untuk data training.	Memiliki nilai akurasi 57.14% untuk data training.

Berdasarkan tabel diatas dapat diketahui bahwa nilai akurasi tertinggi berdasarkan rapidminer ialah algoritma naives bayes sejumlah 78.10% sedangkan random forest 57.14%. Sehingga peneliti menyarankan implementasi development kedepanya dapat menggunakan algoritma naives bayes.

Hal tersebut dikarenakan nilai keakuratan naives bayes cukup tinggi. Nilai akurasi tertinggi merupakan nilai yang dapat digunakan sebagai acuan keakuratan penelitian [15].

4. KESIMPULAN

Setelah melakukan analisis, perancangan, dan pengujian maka dapat diperoleh kesimpulan bahwa evaluasi pembahasan sebelumnya dan dari pengujian yang telah dilakukan, penerapan data mining klasifikasi nasabah pada Bank Palembang menghasilkan akurasi untuk naives bayes 78,10% dan random forest 57,14% dan tujuan penelitian telah dijawab dengan nilai akurasi tersebut. Atribut yang digunakan pada naives bayes dan random forest merupakan atribut yang sama label, angsuran, pendapatan, jenis kelamin dan pekerjaan.

REFERENSI

- [1] Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Klasifikasi Penerimaan Beasiswa. ISSN : 2088-4591.
- [2] Andri, Kunang, Y. N., & Murniati, S. (2013). Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang. Seminar Nasional Informatika , A1-A2.
- [3] Dewi, S. (2019). Komparasi Metode Algoritma Data Mining Pada Prediksi Uji Kelayakan Credit Approval Pada Calon Nasabah Kredit Perbankan. Jurnal Khatulistiwa Informatika, Vol VII, No. 1, ISSN : 2579 - 633X59.
- [4] Fiastantyo, G. (2015). Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan Naive Bayes dan Algoritma C4.5 Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. Semarang: <http://eprints.dinus.ac.id>.
- [5] Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). Media Infotama , 130.
- [6] Hastuti, K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif. Seminar Nasional Teknologi Informasi & Komunikasi Terapan (Semantik) , 241-249.
- [7] Hijriana, N., & Muttaqin, R. (2016). Penerapan Metode Decision Tree Algoritma C4.5 Untuk Klasifikasi Mahasiswa Berprestasi. Al Ulum Sains dan Teknologi , 39-40.
- [8] Kusmini, L., & Taufiq, E. (2009). Algoritma Data Mining. Yogyakarta: Andi.
- [9] Moleong, L. J. (1999). Metodologi Penelitian. Bandung: PT. Remaja Rosda Karya.
- [10] Nazir, M. (2003). Metode Penelitian. Jakarta: Ghalia Indonesia, Cet Ke- 5.
- [11] Novianti, B. (2016). Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus SMA Negeri 1 Pontianak). Jurnal Coding, Sistem Komputer Untan ISSN : 2338-493X , 76-77.
- [12] Nugroho, Y. S., & Haryati, M. S. (2015). Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali. Surakarta: KHAZANAH INFORMATIKA | Online ISSN: 2477-698X.
- [13] Prasetyo, E. (2014). Mengelola Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi.
- [14] Wati, A. S. et al.(2022) 'Perbandingan Implementasi Algoritma CT-PRO dan Algoritma C45 Untuk Menentukan Pola Nasabah', (November), pp. 16–17.
- [15] Rudiansyah, Rudiansyah, and Ade S. Wati. "Pencarian Pola Asosiasi Minat Beli Kebutuhan Pokok Menggunakan Perbandingan Algoritma Apriori dan Fp-growth." *Jurnal Ilmiah Betrik*, vol. 12, no. 3, 2021, pp. 299-305.