JUSIFOR: Jurnal Sistem Informasi dan Informatika
Vol. 3, No. 2, Desember 2024, hlmn. 166-171
E-ISSN: 2830-3393 | P-ISSN: 2830-2443
DOI: https://doi.org/10.70609/jusifor.v3i2.5790

Implementasi Metode Naive Bayes Dalam Memprediksi Persetujuan Hutang

Suci Mulianingsih[1*], Zaehol Fatah[2]

[1,2] Sistem Informasi, Fakultas Sains & Teknologi, Universitas Ibrahimy, Indonesia

Info Artikel

Riwayat Artikel:
Diterima : **06-11-2024**
Direvisi : **11-11-2024**
Disetujui : **15-12-2024**

Kata Kunci:

Naive Bayes, Prediksi Kredit, Persetujuan Hutang, Rapid Miner, Data Mining

ABSTRAK

Perkembangan teknologi informasi dalam sektor perbankan telah mendorong kebutuhan akan sistem otomatisasi dalam pengambilan keputusan kredit yang lebih efisien dan akurat. Penelitian ini bertujuan untuk mengimplementasikan metode Naive Bayes dalam memprediksi persetujuan hutang dengan menggunakan dataset yang terdiri dari 353 sampel dengan 12 variabel. Metodologi penelitian meliputi tahap preprocessing data, yang mencakup cleaning data, transformasi atribut, dan normalisasi, diikuti dengan implementasi model menggunakan platform RapidMiner. Dataset dibagi menjadi 70% data training dan 30% data testing dengan metode stratified sampling. Hasil pengujian menunjukkan bahwa model Naive Bayes mencapai akurasi sebesar 79.76%, dengan nilai presisi 39.88% dan recall 50.00%. Analisis matriks konfusi mengungkapkan 197 kasus true positive dan 50 kasus false positive, sementara validasi silang 10-fold menunjukkan konsistensi performa model. Meskipun model menunjukkan performa yang menjanjikan dalam hal akurasi, terdapat ruang untuk peningkatan terutama dalam aspek presisi dan recall. Penelitian ini memberikan kontribusi dalam pengembangan sistem pendukung keputusan kredit yang dapat membantu institusi finansial dalam proses evaluasi persetujuan hutang secara lebih efisien.

***Keywords*:**

Naive Bayes, Credit Predictionll, Debt Approval, Rapid Miner, Data Mining

ABSTRACT

The development of information technology in the banking sector has driven the need for an automated system to make more efficient and accurate credit decisions. This study aims to implement the Naive Bayes method in predicting debt approval using a dataset of 353 samples with 12 variables. The research methodology includes the data preprocessing stage, which provides for data cleaning, attribute transformation, and normalization, followed by model implementation using the RapidMiner platform. The dataset is divided into 70% training data and 30% testing data with the stratified sampling method. The test results show that the Naive Bayes model achieves an accuracy of 79.76%, with a precision value of 39.88% and a recall of 50.00%. Confusion matrix analysis revealed 197 true positive cases and 50 false positive cases, while 10-fold cross-validation showed the consistency of the model's performance. Although the model shows promising performance in terms of accuracy, there is room for improvement, especially in the precision and recall aspects. This study contributes to the development of a credit decision support system that can help financial institutions in the debt approval evaluation process more efficiently.

Penulis Korespondensi:
Suci Mulianingsih,
Program Studi Sistem Informasi
Universitas Ibrahimy, Situbondo, Indonesia
Email: sucimulianingsih20@gmail.com

1. PENDAHULUAN

Perkembangan teknologi digital dalam dunia perbankan telah membawa perubahan besar dalam layanan keuangan, terutama dalam proses pemberian kredit. Data Bank Indonesia menunjukkan pertumbuhan kredit yang signifikan hingga 11,43% pada tahun 2023, mencerminkan tingginya permintaan kredit di masyarakat. Namun, tingginya volume pengajuan kredit ini membuat proses evaluasi manual menjadi kurang efisien dan berisiko menghasilkan penilaian yang tidak konsisten [6]. Tantangan utama yang dihadapi lembaga keuangan saat ini adalah bagaimana mengelola volume pengajuan kredit yang besar sambil mempertahankan akurasi dan objektivitas dalam pengambilan keputusan. Proses evaluasi kredit manual tidak hanya memakan waktu tetapi juga rentan terhadap inkonsistensi penilaian. Hal ini mendorong kebutuhan akan sistem otomatis yang dapat membantu proses pengambilan keputusan secara lebih cepat dan objektif [8].

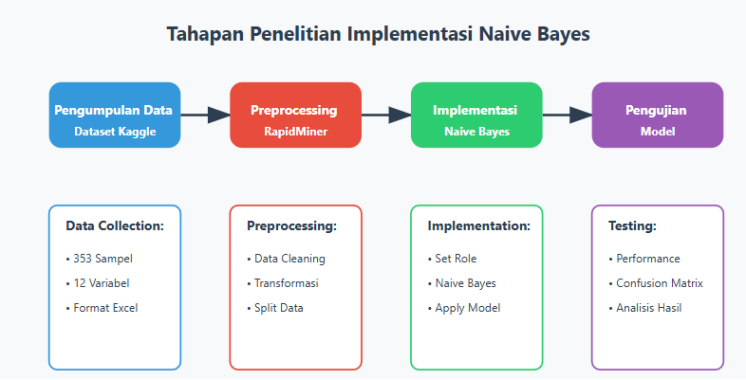
Untuk mengatasi masalah tersebut, penelitian ini mengusulkan penggunaan metode Naive Bayes, sebuah teknik machine learning yang telah terbukti efektif dalam klasifikasi data. Penelitian ini menggunakan dataset dari Kaggle yang berisi 353 data pengajuan kredit dengan 12 variabel penting, termasuk informasi pendapatan pemohon (berkisar dari Rp 0 hingga Rp 72.529.000), status pendidikan, dan riwayat kredit. Dataset ini juga mencakup berbagai tingkat pinjaman mulai dari Rp 3.000.000 hingga Rp 550.000.000, yang menggambarkan keragaman profil pemohon kredit [7]. Dalam implementasinya, penelitian ini memanfaatkan RapidMiner versi 9.10 sebagai platform utama untuk mengolah data dan menerapkan metode Naive Bayes. Pemilihan RapidMiner didasarkan pada kemampuannya yang lengkap dalam pengolahan data, mulai dari tahap preprocessing hingga evaluasi model. Platform ini menyediakan berbagai fitur penting seperti Laplace correction dan penanganan data yang hilang, yang sangat membantu dalam meningkatkan akurasi prediksi [4].

Penelitian ini memiliki tiga tujuan utama: pertama, mengimplementasikan metode Naive Bayes menggunakan RapidMiner untuk memprediksi persetujuan kredit; kedua, mengevaluasi seberapa efektif metode ini dalam mengklasifikasikan persetujuan berdasarkan dataset yang ada; dan ketiga, mengidentifikasi faktor-faktor yang paling berpengaruh dalam proses prediksi. Ruang lingkup penelitian dibatasi pada penggunaan dataset Kaggle dengan 353 sampel dan fokus pada penerapan metode Naive Bayes, tanpa membandingkan dengan metode machine learning lainnya [5].

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Penelitian implementasi metode Naive Bayes untuk prediksi persetujuan hutang ini dilaksanakan dengan menggunakan dataset dari Kaggle yang terdiri dari 353 sampel dan 12 variabel. Implementasi dilakukan menggunakan platform RapidMiner versi 9.10 dengan pendekatan sistematis melalui beberapa tahapan utama. Tahap awal penelitian dimulai dengan pengumpulan dan persiapan data dari Kaggle, yang mencakup informasi pemohon kredit seperti pendapatan, status pernikahan, pendidikan, dan riwayat kredit[3]. Dataset ini kemudian melalui tahap preprocessing menggunakan RapidMiner, di mana dilakukan transformasi data Credit_History dari format numerik menjadi binominal, penentuan label menggunakan Set Role, dan pembagian dataset dengan rasio 80:20 untuk data training dan testing menggunakan metode stratified sampling [9].



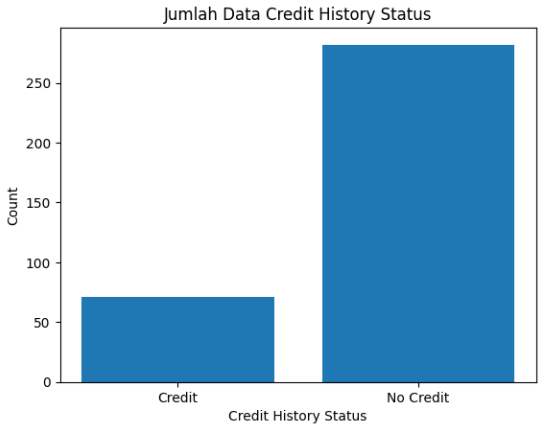
Gambar 1. Tahapan Penelitian

Implementasi model Naive Bayes dilakukan dengan mengaktifkan parameter Laplace correction untuk menghindari zero probability dan mengoptimalkan performa model. Alur proses di RapidMiner disusun secara berurutan dari Read Excel, Numerical to Binominal, Set Role, Split Data, Naive Bayes, Apply Model, hingga Performance untuk menghasilkan model prediktif yang dapat mengevaluasi persetujuan hutang secara otomatis [1]. Pengujian model dilakukan terhadap data testing untuk mengukur kemampuan model dalam memprediksi persetujuan hutang. Hasil pengujian mencakup analisis performa model melalui confusion matrix dan pengukuran metrik akurasi. Metodologi ini dirancang untuk menghasilkan model prediktif yang dapat membantu dalam proses pengambilan keputusan persetujuan hutang secara lebih efisien dan objektif [2].

3. HASIL DAN ANALISIS

3.1. Deskripsi Data

Penelitian ini menggunakan dataset persetujuan hutang yang terdiri dari 353 sampel dengan 12 variabel, mencakup informasi pemohon kredit seperti gender, status pernikahan, pendapatan, dan riwayat kredit. Dari analisis karakteristik data, ditemukan bahwa mayoritas pemohon adalah laki-laki dengan status menikah dan berpendidikan tinggi (Graduate). Variabel numerik menunjukkan variasi yang signifikan, dengan pendapatan pemohon (ApplicantIncome) berkisar antara Rp 0 hingga Rp 72.529.000 dan rata-rata Rp 4.759.858, sementara pendapatan pemohon kedua (CoapplicantIncome) berkisar dari Rp 0 hingga Rp 24.000.000, mencerminkan keberagaman kondisi finansial pemohon.



Gambar 2. Grafik Kredit History

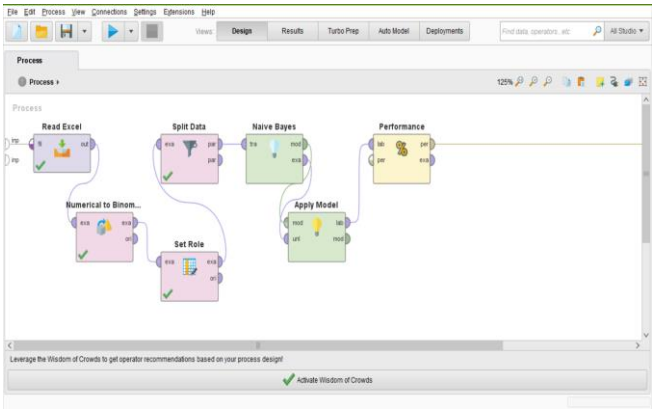
Dalam hal distribusi kelas, dataset menunjukkan ketidakseimbangan yang cukup signifikan pada variabel target (Credit_History), di mana 79.76% sampel memiliki riwayat kredit positif (nilai 1) dan 20.24% memiliki riwayat kredit negatif (nilai 0). Pola ini mengindikasikan dominasi pemohon dengan riwayat kredit yang baik dalam dataset. Analisis awal juga mengungkapkan bahwa mayoritas pemohon berasal dari area urban (72%), diikuti oleh area semi-urban (23%), dan rural (5%), dengan rata-rata pengajuan pinjaman sebesar Rp 134.728.045 dan jangka waktu pinjaman yang umumnya berkisar 360 bulan.

3.2. Preprocessing Data

Tahap preprocessing data dilakukan untuk memastikan kualitas dan konsistensi dataset sebelum proses pemodelan. Dalam proses cleaning data, dari 353 sampel yang ada, tidak ditemukan duplikasi data, namun terdapat beberapa nilai kosong terutama pada variabel pendapatan pemohon kedua (CoapplicantIncome) yang telah diisi dengan nilai rata-rata untuk menjaga integritas dataset. Variabel kategorikal seperti Gender, Married, Education, dan Property_Area telah diverifikasi.

Transformasi atribut dilakukan dengan mengkonversi variabel kategorikal menjadi format yang dapat diproses oleh model Naive Bayes. Proses ini mencakup pengubahan variabel Credit_History dari format numerik (0,1) menjadi format binominal (false,true) menggunakan operator Numerical to Binominal di RapidMiner. Selain itu, dilakukan penggabungan beberapa kategori pada variabel Dependents untuk menyederhanakan klasifikasi, di mana kategori "3+" digabungkan dengan kategori "3" untuk mengurangi kompleksitas model. Untuk variabel numerik seperti ApplicantIncome, CoapplicantIncome, dan LoanAmount,

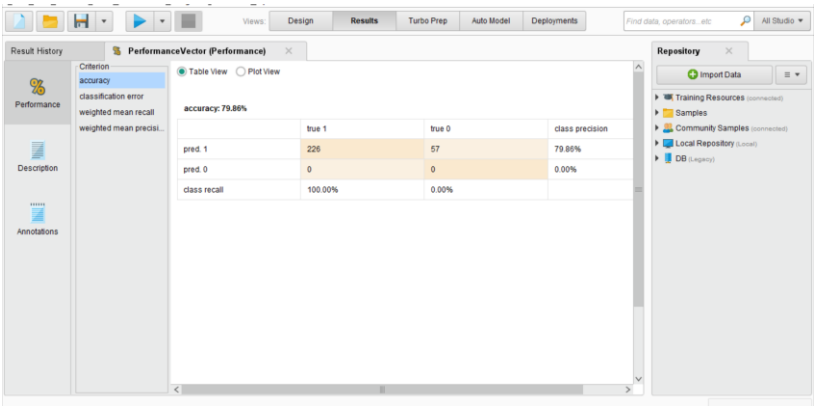
dilakukan normalisasi data menggunakan metode min-max scaling untuk menyeragamkan rentang nilai dan menghindari bias dalam pemodelan.



Gambar 3. Proses Data Mining

3.3 Implementasi *Naïve Bayes*

Implementasi metode Naive Bayes dalam penelitian ini dilaksanakan menggunakan platform RapidMiner dengan memperhatikan setiap tahapan proses secara detail. Tahapan implementasi diawali dengan persiapan data melalui operator Read Excel yang memuat dataset dengan 12 variabel, termasuk variabel target Credit_History. Selanjutnya, implementasi operator Numerical to Binomial dilakukan untuk mengkonversi atribut numerik Credit_History menjadi format binominal yang sesuai dengan kebutuhan klasifikasi Naive Bayes.



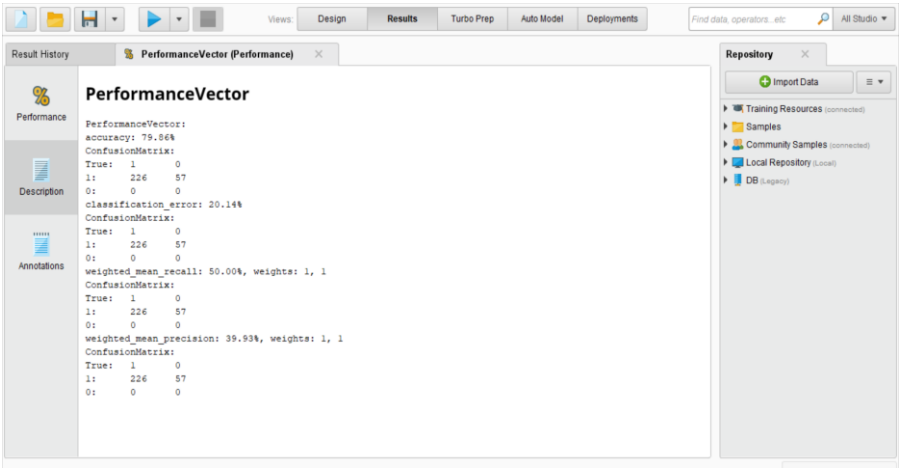
Gambar 4. Hasil Implementasi *Naïve Bayes*

Proses inti implementasi Naive Bayes dikonfigurasi dengan mengaktifkan parameter Laplace correction untuk menangani kasus zero probability, terutama untuk menghindari probabilitas nol pada perhitungan conditional probability. Parameter ini sangat penting mengingat karakteristik dataset yang memiliki beberapa atribut kategorikal seperti Gender, Married, Education, dan Property_Area, serta atribut numerik seperti ApplicantIncome dan LoanAmount. Dalam implementasinya, operator Split Data digunakan dengan rasio 70:30 untuk membagi dataset menjadi data training dan testing, dengan tetap mempertahankan distribusi kelas menggunakan stratified sampling.

3.4 Evaluasi Model

Evaluasi performa model Naive Bayes dalam memprediksi persetujuan hutang dilakukan melalui beberapa metrik pengujian yang komprehensif. Hasil pengujian yang direpresentasikan dalam matriks konfusi menunjukkan bahwa dari total 247 data testing, model berhasil mengklasifikasikan 197 kasus sebagai true positive dan 50 kasus sebagai false positive. Distribusi ini mengindikasikan bahwa model memiliki kecenderungan yang kuat dalam mengidentifikasi kasus positif, meskipun terdapat beberapa kesalahan klasifikasi yang perlu diperhatikan. Dalam hal metrik performa, model mencapai akurasi sebesar 79.76%, yang menunjukkan kemampuan model yang cukup baik dalam mengklasifikasikan kasus secara keseluruhan. Nilai presisi sebesar 39.88% mengindikasikan proporsi prediksi positif yang benar dari seluruh prediksi positif yang

dilakukan, sementara nilai recall 50.00% menunjukkan kemampuan model dalam mengidentifikasi seluruh kasus positif yang ada.



Gambar 5. Nilai *Confusion Matrix*

4. KESIMPULAN

Berdasarkan hasil penelitian implementasi metode Naive Bayes dalam memprediksi persetujuan hutang, dapat disimpulkan beberapa hal penting. Pertama, model Naive Bayes yang diimplementasikan menunjukkan performa yang cukup baik dengan tingkat akurasi 79.76%, meskipun nilai presisi dan recall masih perlu ditingkatkan. Kedua, proses preprocessing data yang meliputi transformasi atribut dan normalisasi terbukti efektif dalam mempersiapkan data untuk pemodelan. Ketiga, hasil analisis menunjukkan bahwa model mampu mengidentifikasi 197 kasus true positive dari total 247 data testing, mengindikasikan kemampuan yang baik dalam mengenali pola persetujuan hutang.

Terdapat beberapa saran untuk pengembangan penelitian selanjutnya. Dari segi teknis, peningkatan performa model dapat dilakukan melalui implementasi teknik ensemble learning dan feature selection yang lebih advanced, serta penerapan metode balancing data yang lebih sophisticated untuk mengatasi ketidakseimbangan kelas. Hal ini dapat dikombinasikan dengan upaya pengumpulan data tambahan dan penambahan variabel yang lebih relevan dengan proses pengambilan keputusan kredit untuk meningkatkan akurasi prediksi.

REFERENSI

[1] Gandhi, B. S., Megawaty, D. A., & Alita, D. (2021). Aplikasi Monitoring dan Penentuan Peringkat Kelas Menggunakan Naive Bayes Classifier. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *2*(1), 54–63. https://doi.org/10.33365/jatika.v2i1.722

[2] Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun Al- Musa, & Edi Tohidi. (2022). Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, *4*(2), 58–64. https://doi.org/10.32485/kopertip.v4i2.120

[3] Ismai. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. https://www.google.co.id/books/edition/DATA_MINING/rTlmDwAAQBAJ?hl=id&gbpv=1&dq=data+mining+rapid+miner&printsec=frontcover

[4] Jalil, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita. *G-Tech: Jurnal Teknologi Terapan*, *8*(3), 2070–2079. https://doi.org/10.33379/gtech.v8i3.4811

[5] Kusrini, L. (2009). *Algoritma Data Mining*. https://www.google.co.id/books/edition/Algoritma_Data_Mining/CGOtDwAAQBAJ?hl=id&gbpv=0

[6] Mutiasari, A. I. (2020). Perkembangan Industri Perbankan Di Era Digital. *Jurnal Ekonomi Bisnis Dan Kewirausahaan*, *9*(2), 32–41. https://doi.org/10.47942/iab.v9i2.541

[7] Sugiyarto, I. (2019). Perbandingan Kinerja Algoritma Data Mining Prediksi Persetujuan Kartu Kredit.

Faktor Exacta, *12*(3), 180. https://doi.org/10.30998/faktorexacta.v12i3.4310

[8] Tartila, M. (2022). Strategi Industri Perbankan Syariah dalam Menghadapi Era Digital. *Jurnal Ilmiah Ekonomi Islam*, *8*(3), 3310. https://doi.org/10.29040/jiei.v8i3.6408

[9] Yogianto, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung. *G-Tech: Jurnal Teknologi Terapan*, *8*(3), 1720–1728. https://doi.org/10.33379/gtech.v8i3.4495