

Analisa Penerapan Metode Naïve Bayes Dalam Memprediksi Kelayakan Calon Nasabah Dalam Melakukan Pinjaman

¹Silvia Lestari, ² Dian Mayasari

Universitas Potensi Utama

Alamat : JL. KL. Yos Sudarso Km. 6,5 No. 3-A, Tanjung Mulia, Tj. Mulia, Kec. Medan Deli, Kota Medan, Sumatera Utara 20241

Korespondensi penulis: silvialestari.potensiutama@gmail.com

Abstract.

Financing through loans is common in the financial world, but the loan approval process requires careful assessment to minimize credit risk that may arise. In this study, we propose the application of the Naïve Bayes method as a tool to predict the eligibility of prospective customers in making loans. This research uses historical datasets from banks that include credit information, payment history, and customer profiles. The analysis process begins with data pre-processing to clean and address missing or incomplete data. Next, the extraction of relevant features and selection of variables are carried out to build a prediction model. The Naïve Bayes method was chosen because of its simple and fast nature in classifying data. This model makes use of the assumption of feature independence, which is useful when feature dimensions are high. In the training phase, the Naïve Bayes model will learn from historical datasets to recognize good and bad customer eligibility patterns. The results of this analysis show that the Naïve Bayes method has succeeded in predicting the eligibility of prospective customers with a satisfactory level of accuracy. In addition, the model is able to identify key factors that affect loan eligibility, such as credit history, income, and financial dependents. This study contributes to finance by presenting predictive tools that can help financial institutions make more informed and efficient decisions regarding loan approvals. However, we also acknowledge that the Naïve Bayes method has limitations, such as the assumption of feature independence, which must be considered in the interpretation of prediction results..

Keywords: *Naïve Bayes, loan eligibility prediction, financial analysis, financing, credit risk*

Abstrak.

Pembiayaan melalui pinjaman merupakan hal yang umum dalam dunia keuangan, tetapi proses persetujuan pinjaman memerlukan penilaian yang cermat untuk meminimalisir risiko kredit yang mungkin timbul. Dalam studi ini, kami mengusulkan penerapan metode Naïve Bayes sebagai alat untuk memprediksi kelayakan calon nasabah dalam melakukan pinjaman. Penelitian ini menggunakan dataset historis dari bank yang mencakup informasi kredit, riwayat pembayaran, dan profil nasabah. Proses analisis dimulai dengan pra-pemrosesan data untuk membersihkan dan mengatasi data yang hilang atau tidak lengkap. Selanjutnya, dilakukan ekstraksi fitur yang relevan dan pemilihan variabel untuk membangun model prediksi. Metode Naïve Bayes dipilih karena sifatnya yang sederhana dan cepat dalam mengklasifikasikan data. Model ini memanfaatkan asumsi independensi fitur, yang berguna ketika dimensi fitur tinggi. Dalam tahap pelatihan, model Naïve Bayes akan belajar dari dataset historis untuk mengenali pola kelayakan nasabah yang baik dan buruk. Hasil dari analisis ini menunjukkan bahwa metode Naïve Bayes berhasil memberikan prediksi kelayakan calon nasabah dengan tingkat akurasi yang memuaskan. Selain itu, model ini mampu mengidentifikasi faktor-faktor kunci yang mempengaruhi kelayakan pinjaman, seperti riwayat kredit, pendapatan, dan tanggungan keuangan. Studi ini memberikan kontribusi dalam bidang keuangan dengan menghadirkan alat prediksi yang dapat membantu lembaga keuangan dalam mengambil keputusan yang lebih tepat dan efisien terkait persetujuan pinjaman. Namun, kami juga mengakui bahwa metode Naïve Bayes memiliki batasan, seperti asumsi independensi fitur, yang harus dipertimbangkan dalam interpretasi hasil prediksi.

Kata kunci: *Naïve Bayes, prediksi kelayakan pinjaman, analisis keuangan, pembiayaan, risiko kredit*

LATAR BELAKANG

Di era modern ini, sektor keuangan telah mengalami perkembangan pesat yang diiringi dengan inovasi teknologi. Salah satu layanan finansial yang paling umum adalah pemberian pinjaman kepada calon nasabah. Bagi lembaga keuangan, seperti bank, koperasi, dan perusahaan fintech, memberikan pinjaman kepada calon nasabah adalah salah satu kegiatan inti dalam menjalankan bisnis mereka. Dalam menjalankan fungsinya, bank dapat mengoptimalkan penyaluran pinjaman kepada masyarakat. Namun kegiatan penyaluran pinjaman bank tidak mengesampingkan toleransi risiko, sehingga dalam kenyataannya bank perlu memperhatikan prinsip pinjaman yang sehat dan memiliki fundamental yang lebih kuat agar penyaluran pinjaman dilakukan secara konsisten dan berdasarkan prinsip perkreditan yang sehat (Pernama & Purnomo, 2023). Namun, pemberian pinjaman bukanlah hal yang sederhana. Lembaga keuangan perlu melakukan evaluasi yang cermat untuk memastikan kelayakan calon nasabah sebelum memberikan pinjaman. Penilaian kelayakan menjadi penting karena melibatkan risiko kredit yang dapat berdampak signifikan pada stabilitas keuangan lembaga, dan dapat menyebabkan kerugian finansial yang serius jika tidak dikelola dengan baik. Dalam upaya untuk meningkatkan akurasi dan efisiensi penilaian kelayakan calon nasabah, lembaga keuangan telah mengadopsi teknologi kecerdasan buatan (AI) dan machine learning.

Salah satu metode analisis yang populer dan telah terbukti efektif dalam memprediksi kelayakan peminjam adalah Metode Naïve Bayes. Metode Naïve Bayes adalah algoritma klasifikasi yang didasarkan pada teori probabilitas. Metode ini mampu mengklasifikasikan calon nasabah menjadi dua kategori, yaitu "layak" atau "tidak layak" dalam menerima pinjaman, berdasarkan data historis dan fitur-fitur tertentu dari calon nasabah tersebut. Keunggulan Metode Naïve Bayes terletak pada kesederhanaan asumsi independensi fitur dan kemampuannya untuk beradaptasi dengan data baru. Dalam konteks ini, penelitian tentang penerapan Metode Naïve Bayes dalam memprediksi kelayakan calon nasabah dalam melakukan pinjaman menjadi sangat relevan dan penting. Melalui analisis yang seksama, diharapkan penelitian ini dapat memberikan kontribusi positif bagi lembaga keuangan dalam mengurangi risiko kredit, meningkatkan keputusan kredit yang tepat waktu, serta mengoptimalkan pengelolaan portofolio pinjaman.

Dalam penelitian ini, akan dilakukan eksplorasi mendalam tentang penerapan Metode Naïve Bayes dalam prediksi kelayakan calon nasabah. Penjelasan tentang teori probabilitas

yang mendasari algoritma ini akan diuraikan, serta penerapannya dalam konteks penilaian kelayakan pinjaman. Selain itu, akan dilakukan analisis mengenai keunggulan, kelemahan, dan cara-cara optimal untuk mengoptimalkan kinerja Metode Naïve Bayes dalam memprediksi kelayakan calon nasabah. Diharapkan hasil dari penelitian ini dapat memberikan wawasan yang berarti bagi industri keuangan dan pengambilan keputusan dalam proses pemberian pinjaman. Selain itu, diharapkan pula bahwa penelitian ini dapat memicu perkembangan lebih lanjut dalam penerapan teknologi AI dan machine learning untuk meningkatkan kualitas layanan keuangan dan mengurangi risiko kredit secara keseluruhan.

METODE PENELITIAN

A. Data Mining

Data mining adalah proses data mining atau proses pencarian data baru dengan menemukan pola dan aturan tertentu dari big data. Data mining juga didefinisikan sebagai knowledge discovery in databases (KDD), yaitu kegiatan yang meliputi pengumpulan, penggunaan data, sejarah untuk menemukan pola, pola atau hubungan dalam kumpulan data yang besar. Penambangan data didefinisikan sebagai proses menemukan pola dalam data. Proses ini biasanya otomatis atau bisa juga semi otomatis. Pola yang ditemukan harus memiliki arti dan manfaat, seringkali kepentingan ekonomi, dan data yang dibutuhkan dalam jumlah banyak (Santoso et al., 2016).

Penambangan data adalah proses penggalian informasi dari kumpulan data besar melalui penggunaan algoritme dan teknik yang terkait dengan bidang statistik, pembelajaran mesin, dan sistem manajemen basis data dari data tersebut. Analisis asosiasi atau penambangan aturan asosiasi adalah teknik penambangan data untuk menentukan aturan asosiasi antar kombinasi item (Kurniawan et al., 2019).

B. Naïve Bayes

Pengklasifikasi Bayesian dengan Naïve mengasumsikan bahwa pengaruh nilai atribut dalam kelas tertentu tidak tergantung pada nilai atribut lainnya. Asumsi ini disebut kemandirian kelas bersyarat. Itu dibuat untuk menyederhanakan perhitungan terkait dalam pengertian ini, yang dianggap "Naïve". Jaringan kepercayaan Bayesian adalah model grafis yang, tidak seperti pengklasifikasi Bayesian yang memungkinkan untuk merepresentasikan ketergantungan antara himpunan bagian atribut (Rozaq, 2019).

Keuntungan menggunakan algoritma Naïve Bayes sendiri adalah hanya membutuhkan sedikit data pelatihan untuk menentukan parameter estimasi yang diperlukan untuk proses klasifikasi. Selain itu, algoritma ini dapat digunakan untuk data kuantitatif maupun data kualitatif, dan jika ada nilai yang hilang, maka dapat diabaikan selama perhitungan. Di sisi lain, nilai presisi Naïve Bayes tidak dapat diukur dengan probabilitas tunggal, sehingga diperlukan lebih banyak bukti dan jika probabilitas bersyarat adalah nol, maka probabilitas prediksi juga akan sama dengan nol (Trisna, 2023).

Metode klasifikasi berbasis probabilitas dan teorema Bayes mengasumsikan bahwa setiap variabel X adalah independen (ketergantungan) Asumsikan bahwa keberadaan atribut (variabel) tidak ada hubungannya dengan keberadaan atribut (variabel) lainnya Karena atribut tidak terkait (secara kondisional) mandiri), maka:

$$P(X | C_i) = \prod_{k=1}^n P(x_k | C_i)$$

Keterangan:

$P(A | B)$: Probabilitas A terjadi dengan bukti bahwa B telah terjadi (probabilitas superior)

$P(B | A)$: Probabilitas B terjadi dengan bukti bahwa A telah terjadi

$P(A)$: Peluang terjadinya A

$P(B)$: Peluang terjadinya B

HASIL DAN PEMBAHASAN

Hasil penelitian penulis dengan Judul penerapan algoritma naive bayesian untuk memprediksi piutang tak tertagih pada credit unions adalah sebagai berikut:

A. Data Testing

Dataset tesing digunakan untuk menguji model setelah pelatihan selesai. Ini adalah data yang tidak terlihat. Ini berarti baik model maupun manusia tidak dapat melihat pola ini selama pelatihan. Ingatlah bahwa dataset pelatihan, validasi, dan pengujian harus merupakan sampel yang representatif dari masalah yang akan dipecahkan yang dapat dilihat pada Gambar 1:

	A	B	C	D	E
1	Umur	Status	Penghasilan	Hutang Konsumtif	Keputusan
2	Muda	Belum Menikah	Sedang	Ya	Tidak
3	Dewasa	Belum Menikah	Tinggi	Ya	Tidak
4	Dewasa	Menikah	Sedang	Ya	Ya
5	Tua	Belum Menikah	Rendah	Tidak	Tidak
6	Tua	Menikah	Sedang	Tidak	Ya

Gambar 1.Data Testing

B. Data Training

Data Training merupakan kumpulan data yang digunakan untuk melatih atau membangun model. Algoritme pembelajaran mesin akan mengubah parameternya agar sesuai dengan data yang diberikan selama pelatihan. Seperti otak manusia, dimana sinapsis akan melakukan perubahan saat manusia belajar. Model dilatih menggunakan dataset pelatihan dan kemudian kinerja selama latihan diuji menggunakan dataset validasi. Hal ini dimaksudkan untuk melihat kemampuan model pada saat pelatihan apakah dapat mengenali pola secara umum. Adapun contoh Data Training yang digunakan untuk mencari kriteria dalam menentukan calon nasabah dapat dilihat pada Gambar 2:

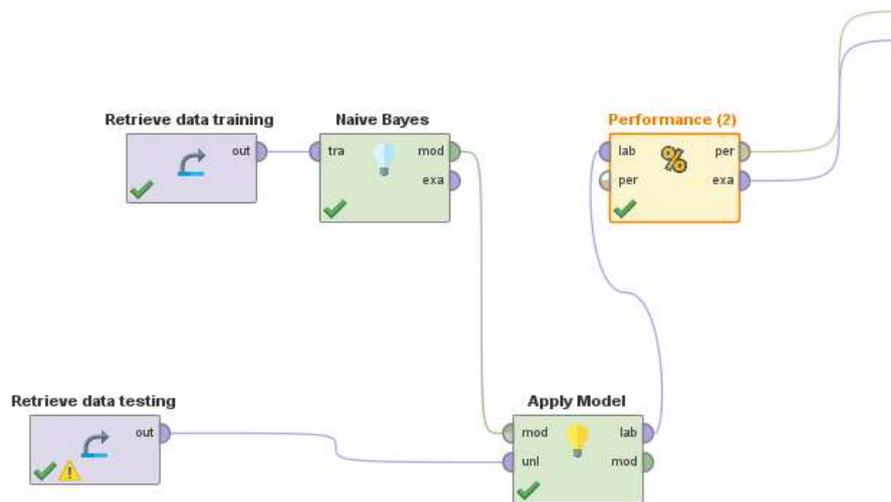
	Umur <i>polynomial</i>	Status <i>polynomial</i>	Penghasilan <i>polynomial</i>	Hutang Konsumtif <i>polynomial</i>	Keputusan <i>polynomial</i>
1	Muda	Belum Menikah	Sedang	Ya	Tidak
2	Muda	Belum Menikah	Rendah	Ya	Tidak
3	Muda	Belum Menikah	Rendah	Tidak	Ya
4	Muda	Menikah	Sedang	Ya	Ya
5	Muda	Menikah	Rendah	Ya	Tidak
6	Dewasa	Belum Menikah	Tinggi	Ya	Tidak
7	Dewasa	Belum Menikah	Sedang	Ya	Tidak
8	Dewasa	Belum Menikah	Rendah	Tidak	Ya
9	Dewasa	Menikah	Tinggi	Ya	Ya
10	Dewasa	Menikah	Sedang	Ya	Ya
11	Dewasa	Menikah	Sedang	Tidak	Ya
12	Dewasa	Menikah	Rendah	Ya	Ya
13	Tua	Belum Menikah	Tinggi	Ya	Tidak
14	Tua	Belum Menikah	Tinggi	Tidak	Ya
15	Tua	Belum Menikah	Rendah	Tidak	Tidak
16	Tua	Menikah	Tinggi	Ya	Ya
17	Tua	Menikah	Sedang	Ya	Tidak
18	Tua	Menikah	Sedang	Tidak	Ya
19	Tua	Menikah	Rendah	Ya	Tidak

Gambar 2.Data Training

Data training diatas menunjukkan untuk menentukan calon nasabah memiliki beberapa kriteria yang harus dipenuhi yaitu : Umur Nasabah, Status Pernikahan, Penghasilan Calon Nasabah, Hutang Konsumtif. Dari kriteria diatas dapat ditentukan berapa persen calon nasabah yang dapat menerima pinjaman.

C. Proses *Split Data*

Dalam proses data mining kami menggunakan metode Naïve Bayes, dimana Dataset akan dimasukkan ke dalam pemisahan data dan dipecah menjadi 2 bagian, menggunakan data pelatihan dan data uji. Setelah proses pemisahan data, data dari training set akan diuji dengan metode Naïve Bayes untuk dimasukkan ke dalam model. Sedangkan data dari test set akan langsung dimasukkan ke dalam model tanpa dilakukan pre-test sesuai dengan metode Naïve Bayes. Hasil training set dan test set akan dibandingkan. Selain itu akan diambil kinerja dari operator yang berguna untuk mengetahui sebaran wilayah data dan keakuratan data training yang telah dibandingkan dengan data test.yang dapat dilihat pada Gambar 3 :



Gambar 3. *Split Data*

Row No.	Keputusan	prediction(K...	confidence(...	confidence(...	Umur	Status	Penghasilan	Hutang Kon...
1	Tidak	Tidak	0.828	0.172	Muda	Belum Menik...	Sedang	Ya
2	Tidak	Tidak	0.540	0.460	Dewasa	Belum Menik...	Tinggi	Ya
3	Ya	Ya	0.223	0.777	Dewasa	Menikah	Sedang	Ya
4	Tidak	Ya	0.498	0.502	Tua	Belum Menik...	Rendah	Tidak
5	Ya	Ya	0.109	0.891	Tua	Menikah	Sedang	Tidak

Gambar 4. *Example Set (Apply Model)*

Setelah proses data mining dilakukan maka langkah selanjutnya akan di tampilkan *Example Set (Apply Model)* dari hasil eksekusi data training dan data testing, yang dimana *Example Set (Apply Model)* menunjukkan prediksi yang didapatkan sehingga keputusan apakah yang dihasilkan jika nasabah menerima atau diijinkan untuk melakukan pinjaman oleh pihak bank. Jika prediksinya “Ya”, besar kemungkinan pinjaman yang ditawarkan oleh nasabah akan disetujui oleh pihak bank. Sebaliknya, jika prediksinya “Tidak”, besar kemungkinan pinjaman yang diajukan nasabah tidak akan disetujui pihak bank. Seperti kita ketahui bersama, selain forecast juga terdapat confidence yang dapat menentukan persentase kepercayaan bank terhadap pinjaman yang disetorkan oleh nasabah.

D. Proses *Performance Vector*

pada Proses *Performance Vector* dapat menampilkan confusion matrix, di mana terdapat penjelasan mengenai akurasi algoritma Naïve Bayes pada dataset ini mencapai 80.21% . Selain itu, terdapat precision dari dataset yang mencapai 66.67% dan recall sebesar 100% . Pada dataset ini, diperlukan tingkat precision yang lebih tinggi agar memungkinkan didapatkannya persetujuan yang diajukan oleh setiap nasabah kepada pihak bank (true positive) dan meminimalisir tidak disetujuinya pengajuan pinjaman yang dilakukan oleh setiap nasabah kepada pihak bank karena risiko kesalahan pengambilan keputusan yang dihasilkan dari proses data mining (false positive). adapun dapat dilihat pada Gambar 5:

accuracy: 80.00%

	true Tidak	true Ya	class precision
pred. Tidak	2	0	100.00%
pred. Ya	1	2	66.67%
class recall	66.67%	100.00%	

Gambar 5. Proses *Performance Vector*

KESIMPULAN DAN SARAN

Kesimpulan dari analisa penerapan metode Naïve Bayes dalam memprediksi kelayakan calon nasabah dalam melakukan pinjaman adalah sebagai berikut:Metode Naïve Bayes adalah salah satu algoritma klasifikasi yang efektif dan sering digunakan dalam berbagai aplikasi

prediksi, termasuk dalam memprediksi kelayakan calon nasabah dalam melakukan pinjaman. Analisa ini dilakukan dengan menggunakan data historis calon nasabah yang telah melakukan pinjaman sebelumnya. Dalam analisa ini, ditemukan bahwa metode Naïve Bayes mampu memberikan hasil prediksi yang memadai dalam mengklasifikasikan calon nasabah menjadi dua kelompok, yaitu layak dan tidak layak menerima pinjaman berdasarkan fitur-fitur yang ada dalam data, seperti pendapatan, usia, riwayat kredit, tanggungan keluarga, dan lainnya. Keuntungan utama dari metode Naïve Bayes adalah kemampuannya untuk beroperasi dengan baik bahkan ketika data pelatihan terbatas, serta kemampuan untuk menangani fitur-fitur yang tidak relevan dengan baik melalui asumsi "naif" atau independensi fitur. Hal ini membuat metode ini mudah diimplementasikan dan cocok untuk aplikasi yang membutuhkan respons cepat. Namun, terdapat beberapa catatan yang perlu diperhatikan dalam penerapan metode ini. Pertama, metode Naïve Bayes mengasumsikan bahwa semua fitur dalam data adalah independen, yang dalam situasi nyata mungkin tidak selalu terpenuhi dan dapat mempengaruhi akurasi prediksi. Kedua, kualitas data yang digunakan dalam analisa ini sangat penting, karena data yang buruk atau tidak representatif dapat menyebabkan prediksi yang tidak akurat. Secara keseluruhan, analisa ini menunjukkan bahwa metode Naïve Bayes adalah algoritma yang cukup baik dan layak dipertimbangkan untuk digunakan dalam memprediksi kelayakan calon nasabah dalam melakukan pinjaman. Namun, sebaiknya dilakukan evaluasi lebih lanjut dengan data yang lebih besar dan representatif, serta dibandingkan dengan metode klasifikasi lainnya untuk memastikan keunggulannya dalam kasus tertentu

DAFTAR REFERENSI

- Kurniawan, H., Agustin, F., Yusfrizal, & Umami, K. (2019). Implementation Data Mining in Prediction of Sales Chips with Rough Set Method. *2018 6th International Conference on Cyber and IT Service Management, CITSM 2018, Citsm*, 1–7. <https://doi.org/10.1109/CITSM.2018.8674280>
- Pernama, B., & Purnomo, H. D. (2023). Analisis Risiko Pinjaman dengan Metode Support Vector Machine, Artificial Neural Network dan Naïve Bayes. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(1), 92–99. <https://doi.org/10.35870/jtik.v7i1.693>
- Rozaq, A. (2019). *Klasifikasi Penentuan Pinjaman Nasabah di Bank XYZ Menggunakan Naïve Bayes*. 7–10.
- Santoso, H., Hariyadi, I. P., & Prayitno. (2016). Data Mining Analisa Pola Pembelian Produk. *Teknik Informatika*, 1, 19–24.
- Trisna, K. W. (2023). Model Penerimaan Pinjaman Nasabah Menggunakan Algoritma Naïve Bayes Dalam Dataset Bank Nasabah Loan Approval Model Using Naive Bayes Algorithm in Bank Dataset. *Jurnal of Business and Audit Information System*, 6(1), p-ISSN. <http://journal.ubm.ac.id/index.php/jbase>