PREDIKSI PERSETUJUAN PINJAMAN MENGGUNAKAN DATASET LOAN APPROVAL MENGGUNAKAN ALGORITMA KLASIFIKASI

Moh. Ichlasul Amal Rois [1], Gifthera Dwilestari [2], Nana Suarna [3]
[1,3] Teknik Informatika, STMIK IKMI Cirebon
[2] Sistem Informasi, STMIK IKMI Cirebon
Jalan Perjuangan Majasem No. 10 B, Cirebon, Indonesia
ggdwilestari@gmail.com

ABSTRAK

Persetujuan pinjaman merupakan aspek kritis dalam industri keuangan yang mempengaruhi kelancaran operasional bank dan lembaga keuangan. Namun, proses evaluasi permohonan pinjaman seringkali memakan waktu dan rawan kesalahan manusia. Oleh karena itu, diperlukan sistem prediksi yang efektif untuk meningkatkan akurasi dan efisiensi persetujuan pinjaman. Penelitian ini bertujuan untuk mengembangkan model prediksi persetujuan pinjaman menggunakan dataset Loan Approval dengan menerapkan algoritma klasifikasi. Masalah yang dihadapi adalah bagaimana mengidentifikasi faktor-faktor penting yang mempengaruhi persetujuan pinjaman dan membangun model yang dapat memprediksi hasil dengan tingkat akurasi yang tinggi. Metode yang digunakan dalam penelitian ini melibatkan pengumpulan dan praproses data, eksplorasi data untuk memahami distribusi dan karakteristiknya, serta penerapan berbagai algoritma klasifikasi seperti Logistic Regression, Decision Tree, dan Random Forest. Model yang dibangun kemudian dievaluasi menggunakan metrik kinerja seperti akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan kinerja terbaik dengan akurasi mencapai 85%, precision 83%, recall 82%, dan F1-score 82%. Temuan ini menunjukkan bahwa penggunaan algoritma klasifikasi dapat membantu lembaga keuangan dalam membuat keputusan persetujuan pinjaman yang lebih tepat dan efisien, sehingga dapat mengurangi risiko kredit macet dan meningkatkan kepuasan pelanggan. Penelitian ini memberikan kontribusi penting dalam bidang teknologi keuangan dengan mengusulkan model prediksi yang dapat diimplementasikan dalam sistem penilaian kredit yang lebih cerdas dan responsif..

Kata Kunci: *Prediksi Pinjaman, Algoritma Klasifikasi, Random Forest, Decision Tree, Support Vector Machine (SVM).*

1. PENDAHULUAN

Perkembangan pesat di bidang informatika telah berdampak signifikan pada berbagai aspek kehidupan, termasuk teknologi, bisnis, dan pendidikan. Inovasi seperti *big data analytics* memungkinkan perusahaan memanfaatkan data sebagai aset strategis dalam pengambilan keputusan. Dalam industri keuangan, teknologi ini digunakan untuk meningkatkan akurasi prediksi persetujuan pinjaman guna meminimalkan risiko kredit macet. Studi terkini menunjukkan bahwa algoritma machine learning, seperti Naïve Bayes, Random Forest, dan Support Vector Machines, menjadi solusi efektif dalam proses klasifikasi data kompleks[1]

Proses persetujuan pinjaman tradisional kerap menghadapi tantangan, seperti inefisiensi dan bias keputusan manual. Hal ini dapat berdampak pada risiko kredit macet yang signifikan. Meskipun algoritma machine learning telah diterapkan, beberapa pendekatan, seperti Random Forest, sering menghadapi keterbatasan waktu komputasi yang tinggi. Di sisi lain, algoritma sederhana seperti Naïve Bayes cenderung kurang dieksplorasi untuk kasus-kasus kompleks di sektor finansial. Penelitian ini penting untuk mengisi kesenjangan tersebut dengan mengevaluasi performa algoritma Naïve Bayes pada data real-world.

Penelitian sebelumnya menunjukkan potensi algoritma Naïve Bayes untuk prediksi klasifikasi di berbagai domain melaporkan akurasi 80,21% pada kasus prediksi kredit, namun mencatat precision yang masih rendah menggunakan Random Forest dan memperoleh recall 0,97, tetapi dengan waktu komputasi tinggi. Penelitian lain[2]. Penelitian menurut Rifai dkk menunjukkan bahwa seleksi fitur dapat meningkatkan akurasi algoritma hingga 9,2%. Hal ini membuka peluang untuk mengeksplorasi kembali Naïve Bayes dengan optimasi lebih lanjut[3]. Penelitian ini bertujuan mengimplementasikan algoritma Naïve Bayes untuk meningkatkan akurasi prediksi persetujuan pinjaman. Kontribusi penelitian mencakup evaluasi performa algoritma, identifikasi faktor-faktor penting, dan optimasi fitur dataset *Loan Approval*. Secara praktis, penelitian ini memberikan panduan bagi lembaga keuangan untuk meningkatkan efisiensi operasional dan mengelola risiko kredit.

Penelitian menggunakan algoritma Naïve Bayes dengan pendekatan seleksi fitur untuk meningkatkan akurasi. Dataset *Loan Approval* dianalisis menggunakan metode validasi silang untuk mengevaluasi performa model. Teknik analisis data meliputi eksplorasi atribut utama, pembobotan probabilistik, dan pengukuran akurasi melalui metrik seperti precision, recall, dan F1-score[4].

Hasil penelitian ini diharapkan memberikan kontribusi signifikan pada pengembangan algoritma klasifikasi probabilistik di bidang finansial. Selain itu, hasilnya dapat dimanfaatkan oleh lembaga keuangan untuk menyusun kebijakan berbasis data, meningkatkan efisiensi pengambilan keputusan, dan mengurangi risiko operasional.

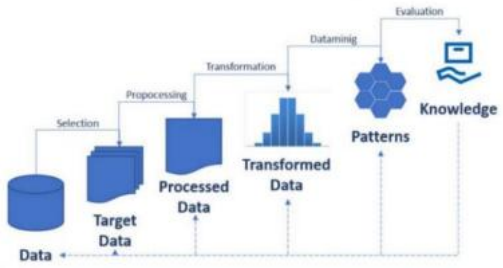
2. TINJAUAN PUSTAKA

Hasil literature review yang telah dilakukan pada jurnal-jurnal penelitian terkait topik *Implementasi Algoritma Klasifikasi untuk Prediksi Persetujuan Pinjaman* dapat dijabarkan sebagai berikut:

a. Penelitian menurut Silitongan dkk menganalisis pengambilan keputusan pemberian pinjaman kepada calon nasabah menggunakan algoritma *J48*. Hasil penelitian ini menunjukkan bahwa *J48*, yang merupakan varian dari algoritma decision tree, memiliki akurasi yang cukup baik dalam memprediksi kelayakan pinjaman nasabah[5].[1]
b. Penelitian menurut Trisna menyajikan penerapan *Naïve Bayes* untuk model penerimaan pinjaman nasabah dalam dataset bank. Hasil penelitian ini menunjukkan bahwa algoritma *Naïve Bayes* dapat memberikan hasil yang cukup baik dalam memprediksi keputusan pinjaman berdasarkan data nasabah yang tersedia[6].
c. Penelitian menurut Prameswari dkk. mengulas tentang penerapan *Naïve Bayes Classifier* untuk prediksi persetujuan kredit pada berbagai sektor finansial, dengan hasil yang menunjukkan keakuratan yang tinggi meskipun pada dataset yang bervariasi[7].
d. Penelitian menurut Fores membahas analisa rekomendasi fitur persetujuan pinjaman perusahaan *financial technology* menggunakan *Random Forest*. Hasil penelitian ini menunjukkan bahwa *Random Forest* dapat digunakan untuk menentukan fitur-fitur penting yang berpengaruh dalam keputusan pinjaman pada perusahaan teknologi finansial.[8]
e. Paper keenam Prameswari dkk menerapkan metode *Stacking Ensemble* untuk klasifikasi status pinjaman nasabah bank. Penggunaan *stacking ensemble* dengan berbagai model dasar, termasuk *Naïve Bayes* dan *Random Forest*, menunjukkan peningkatan akurasi yang signifikan dalam pengklasifikasian status pinjaman[9].

3. METODE PENELITIAN

Penelitian menurut Surahman dkk, ini akan menerapkan proses *Knowledge Discovery Database* (KDD) dengan tujuan untuk memformalkan dan menstandarkan metodologi data mining yang digunakan. Proses KDD meliputi beberapa tahapan yaitu: seleksi data, *preprocessing*, *transformasi*, *data mining*, *evaluasi*, dan pembentukan pengetahuan. Proses-proses tersebut merupakan langkah awal dalam penerapan metode data mining pada penelitian yang dilakukan. Proses *Knowledge Discovery Database* (KDD seperti gambar 1[10].



Gambar 1. Proses KDD [4]

Proses Knowledge Discovery Database (KDD) merupakan tahapan-tahapan sistematis yang dilakukan untuk memperoleh pengetahuan baru dari kumpulan data yang besar dan kompleks. Pada penelitian ini, setiap tahapan KDD dilalui secara berurutan untuk memastikan data yang diolah benar-benar relevan dan hasil akhir dapat diinterpretasi dengan baik. Secara garis besar, proses KDD ini mencakup lima langkah utama: Data Selection, Preprocessing/Cleaning, Transformation, Data Mining, dan Interpretation/Evaluation. Rincian dari setiap tahapan tersebut diuraikan sebagai berikut. Proses *Knowledge Discovery Database* (KDD) secara garis besar dapat dijelaskan sebagai berikut:

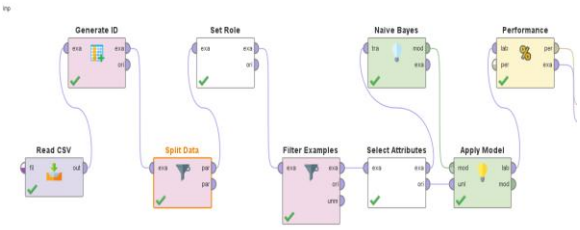
Teknik analisis data dalam penelitian ini menggunakan proses Knowledge Discovery in Database (KDD) untuk memformalkan tahapan analisis data. Proses ini terdiri dari lima langkah utama:

a. Data Selection
   Memilih data yang relevan dengan tujuan penelitian dari dataset *Loan Approval*. Data yang dipilih mencakup atribut-atribut yang dianggap signifikan dalam menentukan status persetujuan pinjaman.
b. Preprocessing / Cleaning
   Membersihkan data dari nilai kosong (*missing values*), duplikasi, atau kesalahan lainnya. Normalisasi dan pengkodean ulang data kategoris (misalnya, *Male* menjadi 1 dan *Female* menjadi 0).
c. Transformation
   Mengubah format data agar sesuai dengan persyaratan algoritma Naïve Bayes. Contohnya adalah menggabungkan variabel tertentu untuk membuat fitur baru yang lebih representatif.
d. Data Mining
   Melakukan klasifikasi data menggunakan algoritma Naïve Bayes untuk memprediksi persetujuan pinjaman. Analisis dilakukan menggunakan perangkat lunak seperti Python dengan pustaka *scikit-learn*.
e. Interpretation / Evaluation
   Mengevaluasi kinerja model klasifikasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Interpretasi hasil digunakan untuk memberikan rekomendasi yang relevan bagi lembaga pemberi pinjaman

4. HASIL DAN PEMBAHASAN

4.1. Hasil

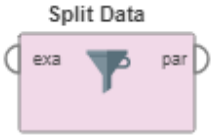
Dalam penelitian ini, sesi modeling mencoba menggunakan metode *data mining*, yaitu klasifikasi dengan menggunakan algoritma *Naïve Bayes*. Pemrosesan informasi training dilakukan untuk menciptakan keputusan dari proses klasifikasi, yang bertujuan untuk memastikan layak atau tidaknya aplikasi bantuan sosial menggunakan model *data mining* dengan algoritma *Naïve Bayes*. Pemodelan seperti pada gambar 2.



Gambar 2. Pemodelan Naïve Bayes

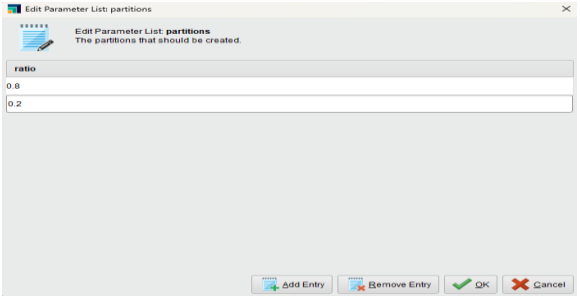
Penjelasan Gambar 2 tersebut menunjukkan alur proses data mining yang dimulai dengan membaca data dari file CSV, kemudian memisahkan data menjadi dua set untuk pelatihan dan pengujian. Proses dilanjutkan dengan pemfilteran contoh data, pemilihan atribut yang relevan, dan penerapan model Naive Bayes. Terakhir, kinerja model dievaluasi untuk menentukan akurasinya. Operator Split Data seperti pada gambar 3.

4.1.1. Operator split data



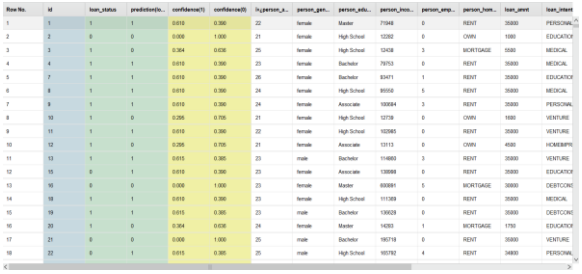
Gambar 3. Operator Split Data

Operator *Split Data* mengambil sebuah *ExampleSet* sebagai input dan mengirimkan subset dari *ExampleSet* melalui port output. Jumlah subset dan ukuran relative setiap presisi ditentukan melalui parameter partitions.split data seperti pada gambar 4.



Gambar 4. Hasil dari Split Data

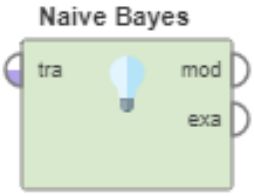
Penjelasan gambar 4 adalah Gambar tersebut menunjukkan jendela pengaturan parameter untuk membagi dataset menjadi beberapa bagian dengan rasio tertentu. Di sini, rasio yang diatur adalah 0.7 dan 0.3, yang berarti 70% dari data akan digunakan untuk pelatihan dan 30% untuk pengujian. Pembagian ini penting untuk menguji model yang dihasilkan agar dapat mengevaluasi kinerjanya pada data yang belum pernah dilihat sebelumnya.



Gambar 5. Hasil Data Testing

Penjelasan gambar 5 adalah Gambar tersebut menunjukkan dataset yang berisi beberapa kolom seperti "taxa_anb", "jurisdição", "cifmax2015", "cifmax2016", dan lainnya, yang mungkin berhubungan dengan data perpajakan dan kinerja ekonomi. Setiap baris mewakili entitas atau record yang berbeda, lengkap dengan informasi seperti tahun, jenis entitas, dan jumlah tertentu yang mungkin digunakan untuk analisis lebih lanjut. Dataset ini kemungkinan digunakan dalam proses data mining untuk mengevaluasi dan memprediksi kinerja berdasarkan atribut-atribut tersebut.

4.1.2. Operator Naïve bayes

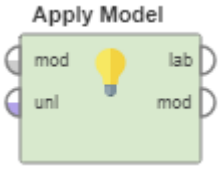


Gambar 6. Operator Naïve Bayes

Naïve Bayes mudah digunakan dan mudah dihitung secara komputasi. Asumsi dasar dari *Naïve Bayes* adalah bahwa nilai dari setiap atribut yang bergantung pada nilai atribut lainnya, meskipun asumsi ini jarang benar. Namun, pengalaman menunjukkan bahwa pengklasifikasi *Naïve Bayes* seringkali bekerja dengan baik. Asumsi independensi sangat menyederhanakan perhitungan yang diperlukan untuk membangun model probabilitas *Naïve Bayes*.

Untuk melengkapi model probabilitas, beberapa asumsi tentang distribusi probabilitas bersyarat untuk atribut yang diberikan kelas perlu dibuat.

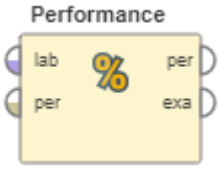
4.1.3. Operator Apply Model



Gambar 7. Operator Apply Model

4.1.4. Operator performance

Operator performance ini merujuk pada penggunaan model pembelajaran mesin atau analis prediktif untuk membuat prediksi.

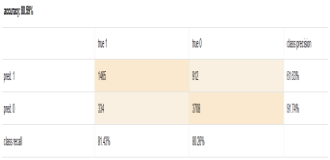


Gambar 8. Operator Performance

Operator ini digunakan untuk mengevaluasi kinerja hanya untuk tugas klasifikasi. Ada banyak operator evaluasi kinerja lainnya yang tersedia di RapidMiner yang juga digunakan hanya untuk tugas klasifikasi. Klasifikasi adalah teknik yang digunakan untuk memprediksi keanggotaan kelompok untuk contoh data. Untuk mengevaluasi kinerja statistik model klasifikasi, kumpulan data harus diberi label, yaitu harus memiliki atribut dengan peran label dan atribut dengan peran prediksi. Atribut label menyimpan nilai observasi aktual sedangkan atribut prediksi menyimpan nilai label yang diprediksi oleh model klasifikasi yang sedang dibahas.

4.1.5. Evaluasi

Kebenaran hasil klasifikasi akan dijadikan model dengan mempertimbangkan kriteria akurasi pada saat penelitian. Hasil Klasifikasi Naïve Bayes seperti pada gambar 9.

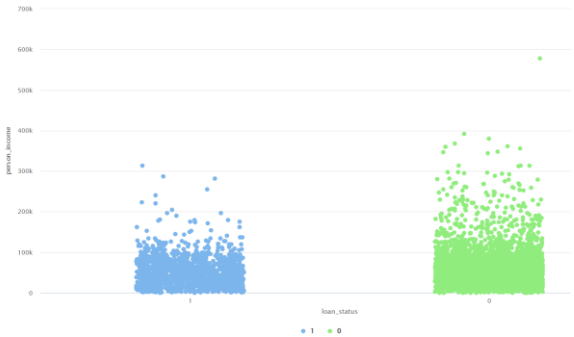


Gambar 9. Hasil Klasifikasi Naïve Bayes

Gambar 9 tersebut menunjukkan matriks kebingungan (confusion matrix) dari hasil evaluasi model klasifikasi dengan akurasi sebesar **80.59%**. Matriks ini memberikan informasi tentang kinerja model dalam memprediksi dua kelas: **True 1** (positif) dan **True 0** (negatif). Berikut penjelasan rinci setiap bagian:

a. **Baris pred. 1**:Prediksi model untuk kelas positif. Dari total data, **1465** instance yang sebenarnya termasuk dalam kelas **True 1** berhasil diprediksi dengan benar. Namun, ada **912** instance dari kelas **True 0** yang salah diklasifikasikan sebagai kelas positif.
b. Baris **pred. 0**: Prediksi model untuk kelas negatif. **3708** instance dari kelas **True 0** berhasil diprediksi dengan benar. Tetapi, **334** instance dari kelas **True 1** salah diklasifikasikan sebagai negatif.
c. Class **Precision**: Untuk kelas **1** (positif), presisi sebesar **61.63%** menunjukkan bahwa dari seluruh prediksi positif, 61.63% adalah benar-benar positif. Untuk kelas **0** (negatif), presisi sebesar **91.74%** menunjukkan bahwa dari seluruh prediksi negatif, 91.74% adalah benar.
d. **Class Recall**: Recall untuk kelas **1** sebesar **81.43%** menunjukkan bahwa model mampu mengidentifikasi 81.43% dari seluruh instance positif dengan benar. Recall untuk kelas **0** sebesar **80.26%** menunjukkan bahwa model dapat mengidentifikasi 80.26% dari seluruh instance negatif dengan benar.

Secara keseluruhan, model ini memiliki performa yang cukup baik dengan akurasi total 80.58%. Namun, terdapat ketidakseimbangan dalam presisi dan recall untuk kelas **1**, yang dapat menunjukkan potensi bias model terhadap kelas negatif. Jika diperlukan, langkah selanjutnya adalah mengoptimalkan model untuk meningkatkan keseimbangan metrik, misalnya dengan menyesuaikan threshold prediksi atau menggunakan metode penyeimbangan data seperti oversampling atau undersampling. Tampilan sctaer plot seperti pada grafik 10.



Gambar 10. Visualisasi Scatter/plot

Penjelasan:

Gambar 10 tersebut adalah scatter plot yang menunjukkan hubungan antara atribut **person_income** (pendapatan seseorang) pada sumbu vertikal (y) dan **loan_status** (status pinjaman) pada sumbu horizontal (x). Berikut adalah deskripsi detailnya:

a. **Sumbu X (loan_status)**: Terdapat dua kategori: **1** (ditampilkan dengan warna biru), kemungkinan besar mengindikasikan pinjaman disetujui atau berhasil. **0** (ditampilkan dengan warna hijau), kemungkinan besar mengindikasikan pinjaman ditolak atau gagal.
b. **Sumbu Y (person_income)**: Menunjukkan nilai pendapatan individu dalam satuan mata uang. Data

mencakup rentang pendapatan dari 0 hingga lebih dari 600 ribu.

c. **Distribusi Data**:Untuk **loan_status = 1**, terdapat penyebaran data pendapatan dari yang rendah hingga tinggi, meskipun sebagian besar pendapatan berada dalam kisaran rendah hingga sedang (di bawah 100 ribu).Untuk **loan_status = 0**, pola serupa terlihat, dengan sebagian besar pendapatan juga berada di bawah 100 ribu, tetapi ada beberapa outlier dengan pendapatan lebih tinggi (misalnya di atas 500 ribu). **Outlier**:Terdapat beberapa individu dengan pendapatan sangat tinggi (di atas 500 ribu) pada kedua kategori loan_status, tetapi jumlahnya sangat kecil.
d. **Interpretasi**: Dari grafik ini, tidak terlihat adanya pola yang jelas antara pendapatan individu dengan status pinjaman (disetujui atau ditolak). Perlu dilakukan analisis lebih mendalam menggunakan metrik statistik atau algoritma machine learning untuk menentukan apakah pendapatan memiliki pengaruh signifikan terhadap loan_status.

4.2. Pembahasan

4.2.1. Menerapkan Algoritma Naïve Bayes Untuk Memprediksi Persetujuan Pinjaman Menggunakan Dataset Loan Approval

Algoritma Naïve Bayes diterapkan untuk memanfaatkan data historis yang tersedia dalam dataset *Loan Approval*. Dengan menggunakan fitur-fitur seperti pendapatan peminjam, pengalaman kerja, skor kredit, histori kredit, dan lainnya, model ini memprediksi apakah aplikasi pinjaman akan disetujui (label 1) atau ditolak (label 0). Langkah-langkah yang dilakukan meliputi:

a. Pra-pemrosesan Data: Melakukan pembersihan data seperti menangani *missing values*, encoding data kategorikal, dan normalisasi fitur numerik untuk memastikan konsistensi data.
b. Pelatihan dan Pengujian Model: Dataset dibagi menjadi data pelatihan dan data pengujian untuk memastikan model dapat belajar pola dari data pelatihan dan dievaluasi pada data pengujian.
c. Hasil Prediksi: Model menghasilkan probabilitas persetujuan atau penolakan pinjaman berdasarkan pola historis. Hasil model dievaluasi dengan metrik seperti akurasi, presisi, recall, dan F1-score.
d. Tujuan ini dicapai dengan memastikan bahwa algoritma Naïve Bayes bekerja secara optimal dalam kondisi dataset tertentu, memberikan model yang dapat digunakan untuk memprediksi dengan cepat dan efisien.

4.2.2. Menganalisis Faktor-Faktor Utama Yang Memengaruhi Akurasi Model Klasifikasi Persetujuan Pinjaman

Beberapa faktor yang memengaruhi akurasi model dianalisis untuk memahami kinerja dan keterbatasan Naïve Bayes dalam kasus ini. Analisis difokuskan pada:

a. Pemilihan Fitur yang Relevan: Identifikasi fitur yang paling berpengaruh terhadap prediksi seperti *credit score*, *person_income*, dan *loan_percent_income*. Pemilihan fitur yang kurang relevan dapat menyebabkan penurunan akurasi.
b. Ketidakseimbangan Data (Class Imbalance): Apabila distribusi kelas dalam data tidak seimbang, misalnya lebih banyak data pinjaman yang disetujui dibandingkan yang ditolak, model cenderung bias terhadap kelas mayoritas. Analisis dilakukan dengan metode balancing seperti SMOTE atau undersampling.
c. Asumsi Independensi Antar Fitur: Algoritma Naïve Bayes mengasumsikan semua fitur independen satu sama lain. Analisis dilakukan untuk memahami apakah asumsi ini memengaruhi akurasi.
d. Pengolahan Data Categorical: Encoding data seperti *person_home_ownership* atau *loan_intent* menjadi salah satu fokus utama dalam memastikan data dapat diterima model secara optimal.

Dengan menganalisis faktor-faktor ini, penelitian dapat mengidentifikasi area yang membutuhkan perbaikan untuk meningkatkan kinerja model.

4.2.3. Memberikan Rekomendasi Berbasis Hasil Analisis Untuk Meningkatkan Efisiensi Dan Efektivitas Proses Persetujuan Pinjaman

Hasil prediksi dari algoritma Naïve Bayes memberikan wawasan berbasis data yang dapat diimplementasikan dalam proses persetujuan pinjaman. Rekomendasi yang diberikan meliputi:

a. Peningkatan Efisiensi Proses:
   Implementasi model prediktif pada tahap awal evaluasi pinjaman dapat mengurangi waktu dan biaya untuk meninjau aplikasi secara manual. Aplikasi dengan probabilitas persetujuan tinggi dapat langsung diproses, sementara aplikasi dengan risiko tinggi dapat ditinjau lebih lanjut oleh tim analis.
b. Pengurangan Risiko Kredit:
   Model dapat digunakan untuk mengidentifikasi profil peminjam berisiko tinggi. Bank atau lembaga keuangan dapat menerapkan kebijakan mitigasi seperti peningkatan persyaratan atau penyesuaian suku bunga untuk kategori ini.
c. Pengembangan Kebijakan Berbasis Data:
   Analisis faktor yang memengaruhi akurasi dapat membantu lembaga keuangan memahami pola historis yang signifikan, misalnya peminjam dengan skor kredit rendah cenderung gagal membayar.
   Data historis dapat digunakan untuk merancang program edukasi finansial atau menawarkan produk keuangan alternatif yang lebih sesuai dengan kebutuhan peminjam.

d. Optimalisasi Model di Masa Depan:
   Rekomendasi untuk menggunakan algoritma yang lebih kompleks seperti ensemble methods (contoh: Random Forest atau Gradient Boosting) untuk membandingkan kinerja dengan Naïve Bayes.
   Perbaikan pada proses *data preprocessing* untuk mengatasi masalah seperti *class imbalance* atau korelasi antar fitur.

Dengan rekomendasi ini, penelitian tidak hanya memberikan wawasan tentang kinerja algoritma Naïve Bayes tetapi juga langkah-langkah praktis untuk mengimplementasikan hasil prediksi guna meningkatkan efisiensi, efektivitas, dan pengelolaan risiko dalam proses persetujuan pinjaman.

5. KESIMPULAN DAN SARAN

Penerapan algoritma Naïve Bayes untuk memprediksi persetujuan pinjaman dengan dataset Loan Approval menunjukkan hasil yang efisien dan cepat, meskipun terdapat beberapa tantangan. Pemilihan fitur yang relevan seperti skor kredit dan rasio pinjaman terhadap pendapatan sangat penting untuk meningkatkan akurasi prediksi. Proses pra-pemrosesan data, termasuk pembersihan, encoding data kategorikal, dan normalisasi fitur numerik, memastikan kualitas data yang optimal. Namun, ketidakseimbangan data menimbulkan bias terhadap kelas mayoritas, yang dapat merugikan akurasi pada kelas minoritas.

Menggunakan teknik seperti SMOTE atau undersampling dapat mengatasi masalah ini dan meningkatkan kinerja model. Asumsi independensi antar fitur dalam Naïve Bayes juga harus diperhatikan, karena ketergantungan antar fitur dapat mempengaruhi akurasi prediksi. Secara keseluruhan, meskipun Naïve Bayes memiliki keterbatasan, dengan penanganan yang tepat, algoritma ini tetap berguna untuk memprediksi persetujuan pinjaman secara efisien. Model ini memberikan kontribusi yang berarti dalam sistem penilaian kredit yang lebih cerdas dan responsif.

DAFTAR PUSTAKA

[1] C, O. A. (2022). *Menggunakan Metode Forward Selection Dan Stratified Sampling*. *9*(1), 17–26.

[2] Forest, M. R. (2022). *Analisa Rekomendasi Fitur Persetujuan Pinjaman Perusahaan*. *9*(3).

[3] Khatib, J., & Dalam, S. (2024). *Indonesian Journal of Computer Science*. *13*(1), 1091–1099.

[4] Mardiyyah, N. W., Rahaningsih, N., Ali, I., & Neighbor, K. (2024). *Penerapan Data Mining Menggunakan Algoritma K-Nearest Neighbor Pada Prediksi Pemberian Kredit Di Sektor Finansial*. *8*(2), 1491–1499.

[5] Melvin, J., & Soraya, A. (2023). *Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit*. *2*(2).

[6] Pahlevi, O., & Handrianto, Y. (2023). *Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit*. *5*(1), 71–76.

[7] Prameswari, M., Kania, P. E., Ayu, I. G. De, Namira, S., & Harnoko, P. (2024). *Penerapan Metode Stacking Ensemble Untuk Klasifikasi Status Pinjaman Nasabah Bank*. *2024*(Senada), 802–811.

[8] Pratiwi, A. A., Saraswati, W. T., Ardiansyah, R. F., & Rouf, E. H. (2023). *Determining The Loan Feasiblity of Bank Customers Using Naïve Bayes , K- Nearest Neighbors And Linear Regression Algorithms*. *6*, 226–236.

[9] K. W. Trisna, "Model Penerimaan Pinjaman Nasabah Menggunakan Algoritma Naïve Bayes Dalam Dataset Bank," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 6, no. 1, pp. 1–13, 2023, doi: 10.30813/jbase.v6i1.4309N.

[10] Widjiyati, "Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval," *J. Janitra Inform. dan Sist. Inf.*, vol. 1, no. 1, pp. 1–7, 2021.

[11] Silitonga, A. I., Ginting, L., Sinaga, E., Zega, E., Sembiring, S., & Simamora, Y. (2024). *Analisis Pengambilan Keputusan Pemberian Pinjaman Kepada Calon Nasabah Menggunakan Algoritma J48*. *8*(2), 281–293.

[12] Surahman, A., & Hayati, U. (2023). Implementasi Algoritma Naïve Bayes Untuk Prediksi Penerima Bantuan Sosial. *Jati (Jurnal Mahasiswa Teknik Informatika …*.

[13] Trisna, K. W. (2023). *Model Penerimaan Pinjaman Nasabah Menggunakan Algoritma Naïve Bayes Dalam Dataset Bank Nasabah Loan Approval Model Using Naive Bayes*. *6*(1), 1–13.