Jurnal Ismetek *ISSN 2406-9841*
Vol. 19 No.1 Juli Tahun 2025 *e-IISN 2986-2973*

KOMPARASI DAN IMPLEMENTASI ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI KREDIT BERMASALAH PADA PT BPR NUSUMMA KLATEN

Sigit Wibisono
Program Studi Teknik Informatika, FTI, Institut Teknologi Budi Utomo Jakarta
wsigitwibisono@gmail.com

Abstrak
Tingkat kredit bermasalah yang tinggi dapat mengganggu stabilitas keuangan lembaga perbankan, sehingga diperlukan sistem klasifikasi yang akurat untuk mendeteksi potensi gagal bayar sejak dini. Penelitian ini bertujuan untuk membangun dan membandingkan model klasifikasi risiko kredit menggunakan algoritma machine learning, yaitu Random Forest, XGBoost, dan Support Vector Machine (SVM). Permasalahan yang diangkat dalam penelitian ini meliputi ketidakakuratan dalam klasifikasi nasabah, kurangnya pemanfaatan data historis, serta belum diterapkannya metode analitik berbasis algoritma cerdas. Metode penelitian mengikuti pendekatan Cross-Industry Standard Process for Data Mining (CRISP-DM) yang mencakup pemahaman bisnis, eksplorasi data, praproses data, pemodelan, evaluasi model, hingga tahap implementasi. Dataset yang digunakan berasal dari laporan historis nasabah kredit di PT BPR Nusumma Klaten. Evaluasi dilakukan dengan mengukur akurasi, precision, recall, F1-score, dan AUC. Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki kinerja terbaik dengan nilai evaluasi yang lebih stabil dibandingkan XGBoost dan SVM. Temuan ini diharapkan dapat membantu lembaga keuangan dalam meningkatkan efisiensi proses analisis risiko kredit dan pengambilan keputusan berbasis data.
Kata kunci : Komparasi, implementasi, algoritma, machine learning

1. PENDAHULUAN

Pemberian kredit oleh pihak Perbankan atau Lembaga Keuangan dengan segala konsekuensi serta resikonya, adalah suatu tantangan utama yang harus dihadapi. Kredit bermasalah atau *Non-performing Loan (NPL)* dapat berdampak pada kesehatan Bank atau suatu Lembaga Keuangan. Banyak pada Bank atau Lembaga Keuangan dalam pemberian kreditnya yang penuh resiko yang ada, proses penentuan tingkat risiko kredit masih dilakukan secara manual yaitu menggunakan pendekatan tradisional. Metode dan prosesnya adalah seperti penelusuran atas riwayat pembayaran calon debitur, bermasaah atau lancar sebagai nasabah atau analisis laporan keuangan dan wawancara dengan calon debitur. Metode ini memiliki keterbatasan karena bergantung pada subjektivitas penilai. Salah satu permasalahan utama dalam pengelolaan kredit adalah bagaimana mengklasifikasikan nasabah ke dalam kategori kredit lancar atau bermasalah berdasarkan riwayat pembayaran. Beberapa nasabah memiliki saldo tabungan tinggi tetapi masih mengalami kesulitan dalam membayar cicilan, sementara yang lain memiliki pinjaman kecil tetapi tetap mengalami gagal bayar

Seiring berkembangnya teknologi *data mining* (Mustika, Yunita, dan Abraham, 2021) dan *machine learning* (Kadek Nonik Erawati, Ni Nengah Dita Ardiani, dan Gede Agus Santiago,2024), maka muncul kebutuhan untuk menerapkan suatu sistem bekerja otomatis yang dapat membantu mengidentifikasi risiko kredit secara lebih akurat dan cepat. Adalah suatu sistem yang diperlukan ini bekerja dengan efektif untuk mengklasifikasikan, kemungkinan nasabah berdasarkan potensi gagal bayar. Dengan meningkatnya volume data keuangan, pemanfaatan metode berbasis kecerdasan buatan seperti *Random Forest* (K. N. Erawati, N. N & D. Ardiani,2024), *XGBoost* dan *Support Vector Machine (SVM)* sebagai solusi yang dapat meningkatkan akurasi dalam memprediksi tantangan risiko kredit. *Random Forest* merupakan teknik *ensemble learning* yang mampu mengurangi *overfitting* dan bekerja dengan baik pada data kompleks. *XGBoost* digunakan untuk mengatasi masalah klasifikasi atau regresi dengan data yang besar dan kompleks. Sementara *SVM* dapat memisahkan kelas data dengan margin maksimal untuk menghasilkan prediksi yang lebih akurat. Ketiga algoritma ini akan digunakan untuk menentukan model terbaik

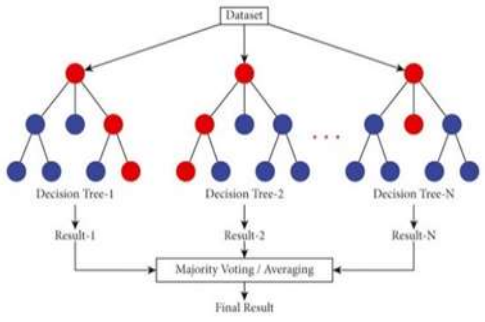
dalam memprediksi nasabah dengan risiko gagal bayar tinggi.

Penerapkan dengan menggunakan pendekatan berbasis *machine learning*, diharapkan bank serta lembaga keuangan dapat meningkatkan efektivitas dalam manajemen risiko kredit. Penggunaan *algoritma Random Fores, XGBoost,* dan *SVM* tidak hanya dapat membantu manajemen dalam mengidentifikasi calon debitur yang berpotensi bermasalah, tetapi juga dapat meningkatkan efisiensi dalam pengambilan keputusan, sehingga mengurangi kemungkinan jumlah kredit bermasalah, yang pada akhirnya mendukung stabilitas keuangan perusahaan dalam jangka panjang.

2. METODOLOGI

2.1 *Random Forest*

Random Forest adalah algoritma *ensemble* yang digunakan untuk klasifikasi dan regresi dengan membangun sejumlah pohon keputusan secara acak dan menggabungkan hasilnya untuk meningkatkan Accuracy. Setiap pohon dibentuk dengan memilih subset acak dari data dan fitur, yang membantu mengurangi *overfitting* dan meningkatkan kemampuan generalisasi. Prediksi akhir diperoleh melalui voting (untuk klasifikasi) atau *averaging* (untuk regresi) dari semua pohon. *Random Forest* dikenal karena keandalannya, kemampuannya mengelola data besar, dan ketahanannya terhadap *overfitting*



Gambar . 1 Arsitektur Algoritma Random Forest (Sumber:https://www.trivusi.web.id/2022/08/algoritma-random-forest_html)

Persamaan : 2.

$\hat{y}$ = mode($\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_\tau$) (2.1)*

Deskripsi:

$\hat{y}$: prediksi yang diberikan oleh pohon keputusan ke-t.

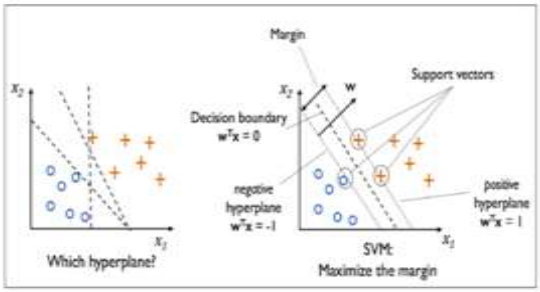
τ : jumlah pohon dalam hutan.

mode :nilai yang paling sering muncul dalam daftar prediksi pohon.

2.2 *Support Vector Machine*

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang diperkenalkan oleh Vapnik pada tahun 1992 untuk keperluan klasifikasi dan regresi. Algoritma ini sangat efektif dalam menangani data non-linear serta data berdimensi tinggi. Prinsip dasarnya adalah menentukan hyperplane paling optimal yang memisahkan dua kelas data dengan memaksimalkan margin, yakni jarak maksimum antara data dari kedua kelompok.

SVM merupakan algoritma pembelajaran mesin yang sering diterapkan dalam tugas klasifikasi dan regresi. Metode ini dirancang untuk menentukan hyperplane terbaik yang memisahkan data dengan margin maksimal. Dalam studi analisis sentimen publik terhadap polusi udara di Jakarta,, *SVM* digunakan untuk mengklasifikasikan kolektabilitas terkait kredit bermasalah berdasarkan data teks yang telah diproses sebelumnya.



Gambar 2. Arsitektur Algoritma SVM (Sumber : https://insancs.medium.com/support-vector-machine-svm-implementation-using-python-4442e9a5babc)

Persamaan : 2.2

$w'x + b = 0$ (2.2)*

Deskripsi:

w : Vektor bobot.

x : Vektor fitur.

b : Bias atau intercept.

2.3 *XGBoost*

Extreme Gradient Boosting (XGBoost) adalah salah satu metode dalam machine learning yang tergolong ke dalam *ensemble method*. *Ensemble method* adalah teknik pembelajaran mesin yang menggabungkan

ISSN 2406-9841
e-IISN 2986-2973

beberapa model dasar untuk menghasilkan prediksi yang lebih akurat dan stabil. *XGBoost* menggunakan salah satu konsep dari *ensemble method* yang disebut dengan *boosting*. Konsep boosting melakukan pembangunan beberapa pohon melalui perbaikan kesalahan yang dilakukan oleh model sebelumnya dengan memberikan bobot lebih pada data yang salah diklasifikasikan.

Menurut Sri Elina Herni Yulianti, Oni Soesanto, dan Yuana Sukmawaty, (2022) Pada metode ini diperlukan fungsi objektif yang berguna untuk menilai seberapa bagus model yang didapatkan sesuai dengan data latih. Karakteristik yang terpenting dari fungsi objektif terdiri dari 2 bagian yaitu nilai pelatihan yang hilang dan nilai regularisasi seperti pada persamaan 2.3 berikut ini.

$$obj\,(\theta) = L(\theta) + \Omega(\theta) \qquad (2.3)^*$$

Dimana L adalah fungsi pelatihan yang hilang, dan Ω adalah fungsi regularisasi, dan θ adalah parameter model terkait. Fungsi pelatihan yang hilang secara umum dapat ditulis seperti pada persamaan 2.4 sebagai berikut.

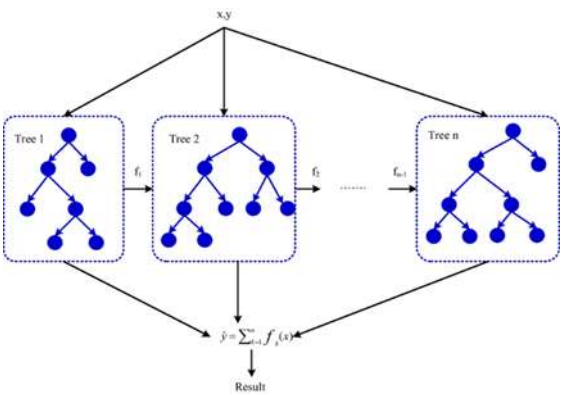
$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) \qquad (2.4)^*$$

Deskripsi :

y_i : adalah nilai data sebenarnya yang dianggap benar

$\hat{y}_i$: adalah hasil nilai prediksi dari model,

n : adalah jumlah iterasi nilai dari mod



Gambar .3 Arsitektur Algoritma XGBoost

(Sumber https://medium.com/@myskill.id/xgboost-fa0a8547e197)

2.4 *CRISP-DM*

*Cross-Industry Standard Process for Data Mining**(CRISP-DM)*** **(**Astri Nur Khusna, 2022), yaitu sebuah standar proses yang umum digunakan dalam praktik data mining lintas industri. Standar ini dikembangkan oleh tiga pelopor utama di bidang data mining, yakni Daimler-Benz, SPSS Inc dan NCR. Penerapan CRISP-DM bertujuan agar proses data mining menjadi sistematis, terukur, dapat diandalkan, serta memenuhi kaidah standar yang berlaku. Secara umum, pendekatan ini terdiri dari enam tahapan utama.

Gambar 4. Diagram Proses *CRISP-DM*
(Sumber : Astri Nur Khusna, 2022)

2.5 Python

Python merupakan salah satu bahasa pemrograman yang sangat populer dan serbaguna dalam menyelesaikan berbagai permasalahan komputasi. Perkembangan terkini menunjukkan bahwa penggunaan Python telah meluas hingga mencakup bidang ekonometrika, statistik, serta analisis numerik secara umum. Dalam hal analisis data, komputasi eksploratif, dan visualisasi interaktif, Python memiliki kapabilitas yang sebanding dengan bahasa pemrograman lainnya seperti R, MATLAB, SAS, dan Stata. Berdasarkan hasil survei terkini, Python menempati posisi teratas sebagai salah satu bahasa pemrograman utama yang digunakan dalam bidang data science maupun pengembangan perangkat lunak secara umum. Keunggulan utama Python terletak pada struktur sintaksis yang sederhana, tingkat keterbacaan kode yang tinggi, serta portabilitasnya yang memudahkan pengembangan lintas platform. (Rommi Kaestria dan Elok Faiqotul Himmah, 2023)

ISSN 2406-9841
e-IISN 2986-2973

[9]. Beberapa pustaka penting yang digunakan dalam analisis data antara lain:

1. ***NumPy*, digunakan untuk untuk** operasi numerik
2. Pandas digunakan untuk manipulasi dan analisis data (D. A. Pamungkas, A. Amali, 2025)
3. Matplotlib digunakan untuk membuat berbagai jenis grafik dengan tingkat kustomisasi yang tinggi
4. Seaborn: Membantu membuat grafik statistik dengan tampilan yang lebih menarik daripada Matplotlib.

Gambar .5 logo Python
(Sumber : https://bpti.uhamka.ac.id/mengenal-python-penjelasan-dan-penggunaannya/)

3. HASIL DAN PEMBAHASAN

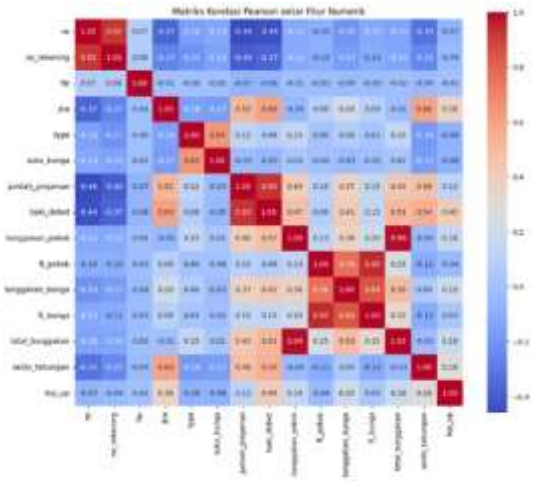
3.1 Strategi Pemecahan Masalah

Untuk menjawab permasalahan yang ada, strategi yang ditawarkan dalam penelitian ini adalah menerapkan metode klasifikasi berbasis *machine learning* guna membangun model prediksi terhadap potensi kredit bermasalah. Strategi ini mencakup beberapa tahapan, adalah:

1. Pengumpulan serta pemahaman terhadap data historis nasabah dari BPR Nusumma Klaten.
2. Praproses data dan eksplorasi fitur yang memiliki relevansi terhadap risiko kredit.
3. Implementasi dan komparasi beberapa algoritma klasifikasi seperti Random Forest, XGBoost, dan SVM.
4. Evaluasi kinerja model menggunakan metrik seperti akurasi, precision, recall, dan F1-score.
5. Rekomendasi model dengan performa terbaik sebagai dasar pengembangan sistem pendukung keputusan dalam proses seleksi kredit, yang selanjutnya diimplementasikan dalam bentuk aplikasi berbasis web menggunakan Streamlit.

3.2 Analisis Korelasi Pearson antar Fitur Numerik

Analisis korelasi digunakan untuk mengetahui tingkat hubungan linear antar variabel numerik dalam dataset. Dalam penelitian ini, korelasi dihitung menggunakan metode Pearson, yang merupakan teknik statistik untuk mengukur kekuatan dan arah hubungan linear antara dua variable



Gambar 6. Matriks Korelasi antar fitur numerik
(Sumber : Dokumentasi penelitian)

Berdasarkan Gambar 4.1 dilakukan analisis korelasi antar variabel numerik guna mengidentifikasi hubungan linier yang kuat maupun lemah antar fitur.

1. Korelasi Positif Kuat
   1) Jumlah_pinjaman sangat berkorelasi dengan :
      - **baki_debet (**0.93**)**
      - tunggakan_pokok (**0.40**)
      - total_tunggakan (**0.43**)
   1) baki_debet memiliki korelasi tinggi dengan :
      - jumlah_pinjaman (**0.93**)
      - total_tunggakan (**0.51**)
   2) t**unggakan_pokok,tunggakan_bunga,** dan total_tunggakan menunjukkan korelasi sangat tinggi antar satu sama lain :
      - tunggakan_pokok dengan total_tunggakan: **0.99**
      - tunggakan_bunga dengan total_tunggakan: **0.50**
      - tunggakan_pokok dengan tunggakan_bunga: **0.36**
      - ft_pokok dengan ft_bunga: **0.95**
2. Korelasi Negatif atau Lemah

1) no, no_rekening, dan hp memiliki korelasi sangat lemah atau mendekati nol terhadap semua fitur — wajar karena sifatnya lebih sebagai pengenal identitas (ID).
2) type dan suku_bunga tidak menunjukkan hubungan yang kuat terhadap fitur numerik lainnya.
3) top_up cenderung memiliki korelasi rendah dengan semua variabel (maksimal sekitar **0.40** dengan baki_debet).

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan mengenai penerapan dan perbandingan algoritma machine learning dalam klasifikasi tingkat kredit bermasalah di PT BPR Nusumma Klaten, dapat disimpulkan beberapa hal sebagai berikut:

1. Berdasarkan hasil evaluasi model, penerapan algoritma machine learning terbukti mampu meningkatkan ketepatan dalam mengklasifikasikan nasabah berdasarkan risiko kredit. Di antara algoritma yang diuji, Random Forest menunjukkan performa paling stabil dan unggul dalam seluruh metrik evaluasi seperti akurasi, presisi, recall, F1-score, dan AUC. Hal ini menunjukkan bahwa pendekatan berbasis machine learning lebih unggul dibanding metode manual dalam proses klasifikasi kredit bermasalah.
2. Data historis nasabah memberikan kontribusi besar dalam pembangunan model prediksi yang akurat. Melalui tahapan data preparation dan analisis mendalam, informasi dari data historis berhasil dimanfaatkan untuk menemukan pola-pola tertentu yang berkaitan dengan kemungkinan gagal bayar. Dengan demikian, pemanfaatan data historis terbukti efektif dalam menunjang sistem klasifikasi risiko kredit berbasis machine learning.
3. Setiap algoritma memiliki keunggulan dan kelemahan masing-masing, namun dalam konteks penelitian ini, Random Forest menjadi pilihan terbaik karena mampu menangani data tidak seimbang dan menghasilkan klasifikasi yang konsisten. Oleh karena itu, pemilihan metode klasifikasi yang sesuai dengan karakteristik data sangat penting untuk mencapai hasil yang optimal dalam prediksi kredit bermasalah.

DAFTAR PUSTAKA

D. A. Pamungkas, A. Amali, and U. D. Soer, 2025. “Analisis sentimen publik terhadap polusi udara di Kota Jakarta: Perbandingan algoritma Support Vector Machine, Naive Bayes, dan Random Forest”, *JUTISI: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 13, no. 3.

K. Erwansyah, B. Andika, and R. Gunawan, 2021. “Implementasi data mining menggunakan asosiasi dengan algoritma Apriori untuk mendapatkan pola rekomendasi belanja produk pada Toko Avis Mobile”, *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 4, no. 1, pp. 148–161.

K. N. Erawati, N. N. D. Ardiani, and G. A. Santiago, 2024. “E-module interaktif berbasis flipbook pada matakuliah machine learning untuk meningkatkan kreatifitas mahasiswa”, *Jurnal Penjaminan Mutu*, vol. 10, no. 1, pp. 45–51

Mustika et al., *Data Mining dan Aplikasinya*, 2021.Bandung: Widina Bhakti Persada Bandung

R. Kaestria and E. F. Himmah, “Implementasi bahasa pemrograman python untuk *path analysis*”, 2023. *Jurnal Komputasi*, vol. 11, no. 2, pp. 105-117.

S. E. H. Yulianti, O. Soesanto, and Y. Sukmawaty,2022.“Penerapan metode Extreme Gradient Boosting (XGBoost) pada klasifikasi nasabah kartu kredit*”, JOMTA: Journal of Mathematics: Theory and Applications*, vol. 4, no. 1, pp. 21–26,.