

Evaluasi Kinerja Algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* untuk Prediksi Risiko Default Nasabah Kartu Kredit

Turlia Indah Sapitri¹, Tia Dwi Anggra Yani², Jelna Anggreni³, Erliyan Redi Susanto⁴

Fakultas Teknik dan Ilmu Komputer, Program Studi Ilmu Komputer

Universitas Teknokrat Indonesia

Turlia_Indah_Sapitri_@teknokrat.ac.id

Abstrak

Prediksi risiko gagal bayar (*default*) nasabah kartu kredit merupakan aspek penting dalam manajemen risiko kredit pada lembaga keuangan. Berbagai penelitian telah menerapkan algoritma klasifikasi untuk prediksi risiko kredit, namun sebagian besar penelitian hanya berfokus pada satu atau dua algoritma sehingga hasil perbandingan performa antar metode masih terbatas. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* dalam memprediksi risiko *default* nasabah kartu kredit. Dataset yang digunakan merupakan data sekunder dari platform Kaggle yang terdiri dari 30.000 data nasabah. Tahapan penelitian meliputi *preprocessing* data, pembagian *training* dan *testing* dengan rasio 80:20, pembangunan model klasifikasi, serta evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Kontribusi penelitian ini terletak pada evaluasi komparatif tiga algoritma klasifikasi dalam satu kerangka eksperimen yang sama menggunakan pendekatan evaluasi multi-metrik. Hasil penelitian menunjukkan bahwa *Random Forest* menghasilkan kinerja terbaik berdasarkan nilai *accuracy* sebesar 0,813500 dan *precision* sebesar 0,630027. Sementara itu, Naive Bayes memperoleh nilai *recall* tertinggi sebesar 0,651181 dan *F1-score* sebesar 0,494363. Temuan penelitian menunjukkan bahwa *Random Forest* lebih unggul dalam ketepatan klasifikasi, sedangkan Naive Bayes lebih efektif dalam mendeteksi nasabah yang berpotensi mengalami gagal bayar.

Kata kunci: *Credit Risk Assessment*, *Kartu Kredit*, *Prediksi Risiko Kredit*, *Naive Bayes*, *Decision Tree*, *Random Forest*

Abstract

Credit card default risk prediction is an important aspect of credit risk management in financial institutions. Although various classification algorithms have been applied to credit risk prediction, most previous studies have focused on only one or two algorithms, resulting in limited comparative insights into their performance. This study aims to evaluate and compare the performance of Naive Bayes, Decision Tree, and Random Forest algorithms in predicting credit card customer default risk. The dataset used in this study is secondary data obtained from Kaggle, consisting of 30,000 customer records. The research process includes data pre-processing, dataset splitting into training and testing sets with an 80:20 ratio, classification model development, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The contribution of this study lies in the comparative evaluation of three classification algorithms within the same experimental framework using a multi-metric evaluation approach. The results indicate that Random Forest achieved the best overall performance with an accuracy of 0.813500 and a precision of 0.630027. Meanwhile, Naive Bayes obtained the highest recall of 0.651181 and the highest F1-score of 0.494363. These findings suggest that Random Forest is more effective in overall classification performance, whereas Naive Bayes is better at identifying customers with a higher likelihood of default.

Keywords: *Credit Risk Assessment*, *Credit Card*, *Credit Risk Prediction*, *Naive Bayes*, *Decision Tree*, *Random Forest*

PENDAHULUAN

Perkembangan penggunaan kartu kredit di sektor perbankan telah memberikan kemudahan bagi masyarakat dalam melakukan transaksi non-tunai serta memenuhi kebutuhan keuangan jangka pendek. Namun, peningkatan penggunaan kartu kredit tersebut juga diikuti oleh meningkatnya risiko gagal bayar (*default*) yang dapat berdampak pada kualitas portofolio kredit dan stabilitas keuangan lembaga perbankan. Risiko gagal bayar umumnya dipengaruhi oleh berbagai faktor, seperti kondisi keuangan nasabah, perilaku penggunaan kartu kredit, serta kemampuan dalam mengelola kewajiban pembayaran. Oleh karena itu, diperlukan suatu pendekatan analitis yang mampu mengidentifikasi potensi risiko kredit secara lebih akurat dan objektif berbasis data historis nasabah (Hazizah & Feranika, 2025). Pemanfaatan teknik *data mining* dan *machine learning* telah banyak diterapkan sebagai salah satu pendekatan dalam pengelolaan risiko kredit karena kemampuannya dalam mengolah data dalam jumlah besar serta menemukan pola-pola tersembunyi yang sulit dianalisis secara konvensional. Dalam praktik *credit scoring* modern, model klasifikasi digunakan untuk membantu lembaga keuangan memprediksi kemungkinan terjadinya gagal bayar sehingga dapat mendukung proses pengambilan keputusan kredit yang lebih efektif dan berbasis data. Berbagai algoritma klasifikasi telah diterapkan dalam prediksi risiko kredit, di antaranya Naive Bayes, *Decision Tree*, dan *Random Forest* (Randy Fachrezi et al., 2025)

Beberapa penelitian terdahulu menunjukkan bahwa algoritma Naive Bayes memiliki keunggulan dari sisi kesederhanaan model dan efisiensi komputasi dalam melakukan prediksi risiko

kredit. Sementara itu, algoritma *Decision Tree* mampu menghasilkan aturan keputusan yang mudah dipahami sehingga banyak digunakan pada sistem pendukung keputusan di sektor keuangan. Di sisi lain, algoritma *Random Forest* sebagai metode *ensemble learning* dilaporkan memiliki tingkat akurasi yang lebih tinggi dan lebih stabil dalam menangani data berukuran besar serta hubungan antarvariabel yang kompleks (Prayesy et al., 2025). Meskipun demikian, hasil penelitian yang diperoleh masih menunjukkan variasi performa karena perbedaan karakteristik *dataset*, jumlah data, distribusi kelas, serta metode evaluasi yang digunakan.

Selain itu, permasalahan yang sering ditemukan dalam prediksi risiko kredit adalah ketidakseimbangan kelas (*class imbalance*), yaitu kondisi ketika jumlah nasabah yang tidak mengalami gagal bayar jauh lebih besar dibandingkan jumlah nasabah yang mengalami gagal bayar. Kondisi ini dapat menyebabkan model klasifikasi cenderung lebih baik dalam memprediksi kelas mayoritas dibandingkan kelas minoritas. Akibatnya, penggunaan metrik *accuracy* saja sering kali tidak cukup untuk menggambarkan kemampuan model dalam mendeteksi nasabah yang benar-benar berisiko gagal bayar. Oleh karena itu, diperlukan evaluasi yang lebih komprehensif menggunakan metrik *precision*, *recall*, dan *F1-score* untuk memperoleh gambaran yang lebih objektif mengenai performa model (Fatah & Khafifah, 2025).

Berdasarkan kajian penelitian terdahulu, sebagian besar penelitian masih berfokus pada penerapan satu algoritma tertentu atau perbandingan terbatas antara dua algoritma klasifikasi. Selain itu, sebagian penelitian hanya menggunakan satu metrik evaluasi utama berupa *accuracy*, sehingga

kemampuan model dalam mendeteksi nasabah berisiko gagal bayar belum dianalisis secara menyeluruh. Dengan demikian, masih terdapat kesenjangan penelitian (*research gap*) terkait evaluasi komparatif algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* dalam satu kerangka eksperimen yang sama menggunakan *dataset* risiko kartu kredit dengan evaluasi multi-metrik.

Kebaruan (*novelty*) penelitian ini terletak pada evaluasi komparatif tiga algoritma klasifikasi, yaitu Naive Bayes, *Decision Tree*, dan *Random Forest*, menggunakan *dataset* risiko *default* kartu kredit yang terdiri dari 30.000 data nasabah dalam satu lingkungan eksperimen yang sama. Penelitian ini tidak hanya menggunakan metrik *accuracy*, tetapi juga menerapkan *precision*, *recall*, dan *F1-score* untuk mengevaluasi kemampuan masing-masing algoritma dalam menghadapi karakteristik data yang tidak seimbang. Selain itu, penelitian ini memberikan analisis komprehensif mengenai keunggulan dan keterbatasan masing-masing algoritma sehingga dapat menjadi dasar dalam pemilihan model yang paling sesuai untuk kebutuhan *credit risk assessment*. Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk mengevaluasi dan membandingkan kinerja algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* dalam memprediksi risiko *default* nasabah kartu kredit. Hasil penelitian diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan model prediksi risiko kredit serta memberikan rekomendasi praktis bagi lembaga keuangan dalam mendukung pengambilan keputusan kredit yang lebih akurat dan berbasis data.

Rumusan Masalah

Rumusan masalah penelitian ini adalah bagaimana perbandingan kinerja algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* dalam klasifikasi risiko kredit berdasarkan berbagai metrik evaluasi.

1. Bagaimana kinerja algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* dalam memprediksi risiko *default* nasabah kartu kredit pada *dataset* yang memiliki karakteristik ketidakseimbangan kelas (*class imbalance*)?
2. Bagaimana perbandingan performa algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score* dalam mendeteksi risiko *default* nasabah kartu kredit?
3. Algoritma manakah yang memberikan performa paling optimal untuk mendukung proses *credit risk assessment* berdasarkan hasil evaluasi multi-metrik?

Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menganalisis kinerja algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* dalam memprediksi risiko *default* nasabah kartu kredit pada *dataset* yang memiliki karakteristik ketidakseimbangan kelas (*class imbalance*).
2. Membandingkan performa algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk memperoleh evaluasi yang lebih komprehensif terhadap kemampuan masing-masing algoritma dalam mendeteksi risiko *default*.
3. Mengidentifikasi algoritma yang memberikan performa paling optimal untuk mendukung proses *credit risk assessment* dan pengambilan keputusan

keputusan kredit berbasis data pada lembaga keuangan.

Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi teoretis dalam pengembangan kajian data mining, khususnya terkait penerapan dan evaluasi kinerja algoritma klasifikasi pada kasus prediksi risiko default nasabah kartu kredit, serta menjadi referensi empiris bagi penelitian selanjutnya di bidang serupa. Selain itu, hasil penelitian ini diharapkan dapat memberikan manfaat praktis bagi lembaga keuangan sebagai bahan pertimbangan dalam memilih algoritma klasifikasi yang paling optimal untuk mendukung pengambilan keputusan dalam manajemen risiko kredit, sehingga dapat meminimalkan potensi terjadinya kredit bermasalah secara lebih efektif dan berbasis data.

Tinjauan Pustaka

Tinjauan pustaka memuat penelitian-penelitian terdahulu yang relevan dengan topik prediksi risiko kredit menggunakan algoritma klasifikasi. Kajian ini bertujuan untuk mengidentifikasi metode yang telah digunakan, hasil yang diperoleh, serta kesenjangan penelitian yang masih dapat dikembangkan pada penelitian saat ini.

Tabel 1. Ringkasan Penelitian Terdahulu

No	Peneliti	Metode	Hasil	Kelemahan
1	(Ariadi, 2020)	Naive Bayes	Akurasi cukup baik dalam prediksi risiko kredit	Hanya menggunakan satu algoritma
2	(Eka Yuni Titik Artaningsih, 2023).	Decision Tree C4.5	Menghasilkan aturan klasifikasi yang	Tidak ada perbandingan dengan metode lain

			mudah dipahami	
3	(Novianti et al., 2023).	Decision Tree C4.5	Mampu mengidentifikasi atribut dominan risiko kredit	Performa bergantung pada dataset
4	(Rahmawati et al., 2025)	Random Forest	Akurasi mencapai 82%	Tidak melakukan studi komparatif

Berdasarkan Tabel 1, penelitian terdahulu menunjukkan bahwa algoritma Naive Bayes, Decision Tree, dan Random Forest memiliki kemampuan yang baik dalam prediksi risiko kredit. Random Forest umumnya menghasilkan tingkat akurasi yang lebih tinggi karena memanfaatkan pendekatan ensemble learning yang mampu mengurangi variansi model. Sementara itu, Decision Tree memiliki keunggulan dari sisi interpretabilitas karena menghasilkan aturan keputusan yang mudah dipahami, sedangkan Naive Bayes menawarkan proses komputasi yang sederhana dan efisien. Meskipun demikian, hasil yang diperoleh masih menunjukkan variasi performa karena perbedaan karakteristik dataset, distribusi kelas, jumlah data, serta metode evaluasi yang digunakan. Selain itu, sebagian besar penelitian masih berfokus pada penggunaan satu algoritma atau perbandingan terbatas antar metode dan hanya menggunakan metrik accuracy sebagai indikator utama evaluasi. Oleh karena itu, masih diperlukan penelitian yang membandingkan algoritma Naive Bayes, Decision Tree, dan Random Forest dalam satu kerangka eksperimen yang sama dengan menggunakan evaluasi multi-metrik untuk memperoleh gambaran performa yang lebih komprehensif.

LANDASAN TEORI

Credit Risk dan Credit Scoring

Risiko kredit (*credit risk*) merupakan risiko yang timbul akibat ketidakmampuan nasabah dalam memenuhi kewajiban pembayaran sesuai perjanjian kredit yang telah disepakati. Oleh karena itu, diperlukan sistem penilaian yang mampu mengidentifikasi potensi gagal bayar secara lebih akurat dan objektif berdasarkan data historis nasabah (M. Rifqi Wirasena, 2024). Menurut penelitian (Kubra Bolukbas, 2025) proses *credit scoring* modern tidak lagi hanya mengandalkan pendekatan statistik tradisional, tetapi juga memanfaatkan algoritma *machine learning* untuk meningkatkan kemampuan prediksi risiko gagal bayar. Pendekatan ini memungkinkan lembaga keuangan melakukan penilaian risiko kredit secara lebih adaptif terhadap karakteristik nasabah yang beragam.

Dalam industri keuangan, proses penilaian risiko kredit umumnya dilakukan melalui sistem *credit scoring*, yaitu metode yang digunakan untuk mengevaluasi tingkat risiko calon debitur berdasarkan karakteristik dan riwayat keuangannya. Sistem ini membantu lembaga keuangan dalam melakukan *credit risk assessment* sehingga keputusan pemberian kredit dapat dilakukan secara lebih konsisten, objektif, dan berbasis data. Perkembangan teknologi informasi juga mendorong penggunaan model prediksi risiko kredit (*default prediction model*) yang memanfaatkan teknik analitik dan *machine learning* untuk meningkatkan akurasi identifikasi nasabah yang berpotensi mengalami gagal bayar.

Perkembangan *credit risk analytics* memungkinkan lembaga keuangan melakukan identifikasi, pengukuran, dan pemantauan risiko kredit secara lebih akurat melalui pemanfaatan data historis pem-

bayaran, karakteristik nasabah, dan pola transaksi. Sejalan dengan berkembangnya konsep *financial machine learning*, berbagai algoritma pembelajaran mesin mulai diterapkan untuk mendukung *credit scoring*, *fraud detection*, *portfolio management*, dan *default prediction*. Penggunaan *machine learning* memungkinkan *preprocessing* data dalam jumlah besar serta identifikasi pola kompleks yang sulit ditemukan menggunakan metode statistik konvensional (Antika et al., 2022); (Ayobami et al., 2023).

Salah satu implementasi penting dari *financial machine learning* adalah pengembangan *default prediction model*, yaitu model yang digunakan untuk memperkirakan kemungkinan seorang nasabah mengalami gagal bayar di masa mendatang. Model ini membantu lembaga keuangan mengurangi potensi kerugian akibat kredit bermasalah serta mendukung pengambilan keputusan kredit secara lebih objektif. Berbagai algoritma klasifikasi telah digunakan dalam pengembangan model prediksi *default*, di antaranya Naive Bayes, Decision Tree, Random Forest, dan Logistic Regression (Gade et al., 2023).

Naive Bayes

Naive Bayes merupakan algoritma klasifikasi dalam *supervised learning* yang bekerja berdasarkan pendekatan probabilistik untuk menentukan kelas suatu data. Algoritma ini memiliki keunggulan berupa kesederhanaan model, efisiensi komputasi, dan kemampuan mengolah data berukuran besar. Namun, Naive Bayes mengasumsikan bahwa setiap atribut bersifat independen, sehingga performanya dapat dipengaruhi oleh adanya korelasi antarvariabel. Dalam bidang *credit risk assessment*, algoritma ini banyak digunakan karena mampu menghasilkan prediksi yang cepat dan efisien. Penelitian oleh (Afandi et

al., 2024) menunjukkan bahwa Naive Bayes memiliki performa yang kompetitif dalam *credit scoring* serta efektif diterapkan pada data kredit dengan jumlah observasi yang besar.

Decision Tree

Algoritma *Decision Tree* mampu membentuk model prediksi yang mudah dipahami karena menghasilkan aturan keputusan dalam bentuk struktur pohon, sehingga dapat digunakan sebagai alat bantu pengambilan keputusan dalam penilaian risiko kredit nasabah. Struktur pohon keputusan terdiri atas *root node*, *internal node*, dan *leaf node* yang merepresentasikan proses pengambilan keputusan berdasarkan atribut tertentu. Kelebihan utama metode ini adalah kemampuannya dalam menghasilkan aturan klasifikasi yang mudah diinterpretasikan. Namun, *Decision Tree* memiliki kelemahan berupa kecenderungan mengalami *overfitting* ketika diterapkan pada data yang kompleks.

Random Forest

Algoritma *Random Forest* merupakan metode *ensemble learning* yang menggabungkan banyak pohon keputusan untuk meningkatkan stabilitas dan akurasi prediksi, terutama dalam menangani data keuangan berdimensi tinggi dan kompleks seperti data risiko kredit (Rizky Saputra Cendikiawan, 2025). Metode ini bekerja dengan membangun sejumlah pohon keputusan secara acak menggunakan teknik *bootstrap sampling* dan pemilihan fitur secara acak. Hasil prediksi akhir diperoleh melalui mekanisme *majority voting*. Pendekatan tersebut membuat *Random Forest* lebih tahan terhadap *overfitting* dan umumnya menghasilkan performa prediksi yang lebih baik dibandingkan model pohon

keputusan tunggal (Anindita Puspa Ayu Prayogi et al., 2025).

Ensemble Learning

Ensemble learning merupakan pendekatan pembelajaran mesin yang menggabungkan beberapa model prediksi untuk menghasilkan performa yang lebih baik dibandingkan model tunggal. Tujuan utama pendekatan ini adalah meningkatkan akurasi, stabilitas, dan kemampuan generalisasi model. Salah satu algoritma yang menerapkan konsep *ensemble learning* adalah *Random Forest*. Dengan menggabungkan banyak pohon keputusan, metode ini mampu mengurangi variansi model dan meningkatkan ketahanan terhadap data yang memiliki karakteristik kompleks maupun mengandung noise.

Evaluation Metrics

Evaluasi kinerja model klasifikasi dilakukan untuk mengukur kemampuan model dalam menghasilkan prediksi yang akurat. Pada penelitian ini digunakan empat metrik evaluasi utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk mengukur tingkat ketepatan prediksi secara keseluruhan. *Precision* digunakan untuk mengukur tingkat ketepatan model dalam mengidentifikasi kelas positif. *Recall* digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh data yang benar-benar termasuk dalam kelas positif. Sementara itu, *F1-score* merupakan rata-rata harmonik antara *precision* dan *recall* yang digunakan untuk memberikan penilaian yang lebih seimbang terhadap performa model.

Hubungan Antar Metode Klasifikasi dalam Prediksi Risiko Kredit

Dalam penelitian prediksi risiko kredit, pemilihan algoritma klasifikasi perlu

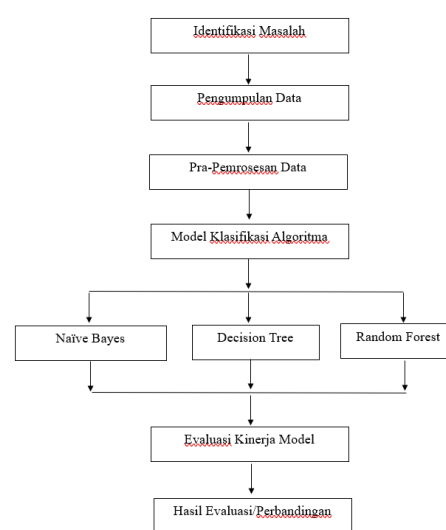
mempertimbangkan karakteristik data dan tujuan analisis. Naive Bayes dipilih karena sederhana dan efisien dalam mengolah data berukuran besar. Decision Tree dipilih karena menghasilkan model yang mudah diinterpretasikan untuk mendukung pengambilan keputusan kredit. Sementara itu, Random Forest dipilih karena merupakan metode *ensemble learning* yang mampu meningkatkan akurasi dan stabilitas prediksi. Ketiga algoritma tersebut memiliki karakteristik yang berbeda, yaitu pendekatan probabilistik pada Naive Bayes, pohon keputusan tunggal pada Decision Tree, dan *ensemble learning* pada Random Forest.

Pada kasus prediksi risiko *default* nasabah kartu kredit yang melibatkan banyak variabel keuangan dan potensi *class imbalance*, ketiga algoritma tersebut menjadi menarik untuk dibandingkan karena mewakili pendekatan klasifikasi yang berbeda. Oleh karena itu, penelitian ini melakukan evaluasi komparatif untuk mengidentifikasi algoritma yang paling sesuai dalam mendukung proses *credit risk assessment* pada lembaga keuangan.

METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen yang menerapkan teknik *data mining* untuk mengevaluasi kinerja algoritma klasifikasi dalam memprediksi risiko *default* nasabah kartu kredit. Algoritma yang digunakan meliputi Naive Bayes, Decision Tree, dan Random Forest yang mewakili pendekatan probabilistik, pohon keputusan, dan *ensemble learning*. Untuk menjaga validitas perbandingan, seluruh algoritma diterapkan pada dataset, tahapan *preprocessing*, dan skenario pembagian data yang sama. Dataset dibagi menjadi 80% *training data* dan 20% *testing data*. Seluruh model dibangun

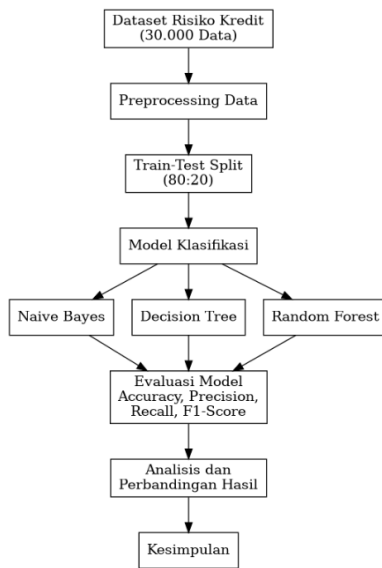
menggunakan parameter standar (*default parameter*) dari pustaka *scikit-learn* sehingga hasil evaluasi mencerminkan karakteristik dasar masing-masing algoritma tanpa proses optimasi parameter. Evaluasi kinerja dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Adapun tahapan penelitian meliputi identifikasi masalah, pengumpulan dan *preprocessing* data, pembangunan model klasifikasi, serta evaluasi dan perbandingan kinerja algoritma untuk menentukan model terbaik.



Gambar 1 Diagram Penelitian

Berdasarkan Gambar 1, penelitian diawali dengan identifikasi masalah terkait prediksi risiko default nasabah kartu kredit, dilanjutkan dengan pengumpulan data dan preprocessing. Data yang telah diproses kemudian digunakan untuk membangun model klasifikasi menggunakan algoritma Naive Bayes, Decision Tree, dan Random Forest. Selanjutnya, kinerja masing-masing model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Tahap akhir dilakukan dengan membandingkan hasil evaluasi untuk menentukan algoritma yang paling optimal dalam mendukung *credit risk assessment*. Selain itu, penelitian ini menggunakan kerangka kerja eksperimen yang

menggambarkan alur pengolahan data, pelatihan model, dan evaluasi kinerja algoritma.



Gambar 2. Experimental Framework Penelitian

ANALISIS DATASET DAN PERANCANGAN MODEL

1. Identifikasi Masalah

Tahap awal dalam penelitian ini adalah identifikasi masalah yang berfokus pada risiko gagal bayar (default) nasabah kartu kredit. Risiko gagal bayar merupakan permasalahan penting dalam manajemen risiko kredit karena dapat berdampak langsung terhadap tingkat kesehatan keuangan lembaga penerbit kartu kredit. Oleh karena itu, diperlukan pendekatan analitis yang mampu memprediksi risiko default nasabah secara lebih akurat berdasarkan data historis.

2. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari platform Kaggle dengan link <https://www.kaggle.com/datasets/laotse/credit-risk-dataset>, yang berisi informasi demografis, riwayat pembayaran, jumlah tagihan, serta data transaksi nasabah kartu kredit yang digunakan untuk memprediksi

risiko gagal bayar (default). Pengolahan data dan pemodelan dilakukan menggunakan Google Colab dengan bahasa pemrograman Python, karena menyediakan lingkungan komputasi berbasis cloud, mendukung pustaka machine learning seperti *scikit-learn*, serta memudahkan proses eksperimen tanpa memerlukan instalasi perangkat lunak secara lokal.

3. Preprocessing Data

Tahap ini meliputi pembersihan data dari nilai kosong (missing values), penyesuaian tipe data numerik dan kategorikal, serta transformasi data agar siap digunakan dalam proses pemodelan. Tahap preprocessing bertujuan untuk meningkatkan kualitas data sehingga hasil klasifikasi menjadi lebih akurat.

Dataset risiko kredit yang digunakan dalam penelitian ini memiliki karakteristik distribusi kelas yang tidak seimbang (imbalanced), di mana jumlah nasabah tidak default lebih dominan dibandingkan nasabah default. Dalam penelitian ini tidak diterapkan teknik penanganan ketidakseimbangan data seperti oversampling atau undersampling, dengan tujuan mengevaluasi kinerja alami masing-masing algoritma pada kondisi data asli. Dampak dari ketidakseimbangan data dianalisis menggunakan metrik recall dan F1-score yang lebih representatif dalam menilai kemampuan model mendeteksi nasabah berisiko gagal bayar.

4. Model Klasifikasi Evaluasi

Model klasifikasi dalam penelitian ini dirancang dengan menerapkan tiga algoritma utama, yaitu Naive Bayes, Decision Tree, dan Random Forest. Naive Bayes digunakan sebagai model probabilistik dengan asumsi independensi antar atribut, Decision Tree digunakan untuk membangun model berbasis aturan yang mudah diinterpretasikan, sedangkan

Random Forest digunakan sebagai metode ensemble untuk meningkatkan stabilitas dan akurasi prediksi. Dataset kemudian dibagi menjadi *training* dan *testing* dengan rasio 80:20, di mana 80% data digunakan untuk proses pelatihan model dan 20% data digunakan untuk pengujian kinerja model.

Seluruh algoritma klasifikasi pada penelitian ini dibangun menggunakan parameter standar (*default parameter*). Pendekatan ini dipilih agar fokus penelitian tetap pada evaluasi dan perbandingan karakteristik kinerja masing-masing algoritma, bukan pada optimasi parameter. Penggunaan parameter standar ini menjadi salah satu keterbatasan penelitian yang dapat dikembangkan pada penelitian selanjutnya. Parameter model yang digunakan pada penelitian ini ditetapkan berdasarkan parameter standar (*default parameter*) dari pustaka *Scikit-learn*. Adapun parameter yang digunakan pada masing-masing algoritma ditunjukkan pada Tabel 2.

Tabel 2. Parameter Model Klasifikasi

Algoritma	Hyperparameter	Nilai
Naive Bayes	<i>var_smoothing</i>	1e-9
Decision Tree	<i>criterion</i>	gini
	<i>max_depth</i>	None
Random Forest	<i>n_estimators</i>	100
	<i>max_depth</i>	None

Parameter model yang digunakan pada penelitian ini ditetapkan berdasarkan parameter standar (*default parameter*) dari pustaka *Scikit-learn*. Penggunaan parameter standar bertujuan untuk menjaga objektivitas perbandingan sehingga hasil evaluasi yang diperoleh lebih merefleksikan karakteristik dasar masing-masing algoritma dibandingkan hasil optimasi parameter tertentu. Dengan demikian, fokus penelitian diarahkan pada evaluasi dan perbandingan performa algoritma Naive Bayes, *Decision Tree*, dan *Random Forest* dalam

memprediksi risiko *default* nasabah kartu kredit.

5. Evaluasi Klasifikasi Model

Metode evaluasi yang digunakan dalam penelitian ini meliputi:

- Accuracy, untuk mengukur tingkat ketepatan prediksi secara keseluruhan
- Precision, untuk mengukur ketepatan prediksi kelas default
- Recall, untuk mengukur kemampuan model mendeteksi nasabah berisiko default
- F1-Score, untuk menyeimbangkan precision dan recall

Penggunaan beberapa metrik evaluasi diperlukan agar kinerja model dapat dinilai secara lebih objektif, terutama pada kasus risiko kredit yang sensitif terhadap kesalahan klasifikasi.

6. Analisis Hasil Evaluasi dan Perbandingan

Hasil evaluasi dari algoritma Naive Bayes, Decision Tree, dan Random Forest dianalisis dan dibandingkan untuk menentukan algoritma dengan performa terbaik. Analisis perbandingan ini dilakukan berdasarkan nilai metrik evaluasi yang diperoleh, sehingga dapat diketahui keunggulan dan kelemahan masing-masing algoritma dalam memprediksi risiko default nasabah kartu kredit.

IMPLEMENTASI DAN PEMBAHASAN Implementasi Eksperimen

Implementasi penelitian ini dilakukan menggunakan Google Colab sebagai lingkungan pengolahan data dengan bahasa pemrograman Python versi 3.11. Dataset risiko kredit nasabah kartu kredit yang digunakan berasal dari platform Kaggle dan terdiri dari 30.000 data nasabah dengan berbagai atribut demografis, riwayat pembayaran, jumlah tagihan, serta data transaksi lainnya. Implementasi penelitian

difokuskan pada proses *preprocessing* data, pembangunan model klasifikasi, dan evaluasi kinerja model sesuai dengan tahapan penelitian yang telah dirancang.

Proses pengolahan data memanfaatkan beberapa pustaka utama, yaitu Pandas untuk pengelolaan data, NumPy untuk komputasi numerik, *Scikit-learn* untuk pembangunan model klasifikasi dan evaluasi, serta Matplotlib untuk visualisasi hasil penelitian. Eksperimen dijalankan pada lingkungan komputasi Google Colab dengan spesifikasi prosesor virtual Intel Xeon dan RAM sekitar 12 GB. Parameter model yang digunakan mengikuti parameter standar sebagaimana dijelaskan pada Tabel 2 sehingga fokus penelitian tetap pada evaluasi karakteristik dasar masing-masing algoritma.

Pada tahap *preprocessing*, dilakukan pemeriksaan kualitas data untuk memastikan data layak digunakan dalam proses pemodelan. Hasil pemeriksaan menunjukkan bahwa seluruh atribut pada dataset tidak memiliki nilai kosong (*missing value*), sehingga tidak diperlukan proses imputasi data. Selanjutnya, data dibersihkan dari kemungkinan duplikasi, dilakukan penyesuaian tipe data, serta transformasi atribut menggunakan metode standarisasi agar setiap fitur memiliki skala yang sebanding. Tahap ini bertujuan untuk meningkatkan kualitas data sehingga dapat menghasilkan model klasifikasi dengan kinerja yang optimal.

Hasil Evaluasi Model

Untuk menganalisis kinerja algoritma klasifikasi secara lebih mendalam, digunakan confusion matrix yang menyajikan jumlah prediksi benar dan salah pada setiap kelas. Confusion matrix digunakan untuk menganalisis kinerja algoritma klasifikasi secara lebih mendalam

dengan menampilkan jumlah prediksi benar dan salah pada setiap kelas.

a. Confusion Matrix Naive Bayes

Tabel 3. Confusion Matrix Algoritma Naive Bayes

Aktual \ Prediksi	Tidak Default (0)	Default (1)
Tidak Default (0)	3.396	1.291
Default (1)	458	855

Berdasarkan Tabel 1, algoritma Naive Bayes mampu mengklasifikasikan dengan benar sebanyak 3.396 data nasabah yang tidak mengalami default dan 855 data nasabah yang mengalami default. Namun, masih terdapat 1.291 data tidak default yang salah diklasifikasikan sebagai default (*false positive*) dan 458 data default yang salah diklasifikasikan sebagai tidak default (*false negative*). Hasil ini menunjukkan bahwa Naive Bayes memiliki kemampuan yang relatif baik dalam mendeteksi nasabah berisiko gagal bayar, yang tercermin dari jumlah *true positive* yang cukup tinggi, sehingga algoritma ini cenderung sensitif terhadap kelas default.

b. Confusion Matrix Decision Tree

Tabel 4. Confusion Matrix Algoritma Decision Tree

Aktual \ Prediksi	Tidak Default (0)	Default (1)
Tidak Default (0)	3.806	881
Default (1)	788	525

Berdasarkan Tabel 2, algoritma Decision Tree berhasil mengklasifikasikan 3.806 data tidak default dan 525 data default secara benar. Namun, terdapat 881 data tidak default yang salah diklasifikasikan sebagai default serta 788 data default yang salah diklasifikasikan sebagai tidak default. Jumlah *false negative* yang relatif tinggi menunjukkan bahwa Decision Tree kurang optimal dalam mendeteksi nasabah yang benar-benar berisiko gagal bayar, sehingga berpotensi menimbulkan risiko kredit jika

digunakan sebagai satu-satunya model prediksi.

c. Confusion Matrix Random Forest

Tabel 5. Confusion Matrix Random Forest

Aktual \ Prediksi	Tidak Default (0)	Default (1)
Tidak Default (0)	4.411	276
Default (1)	843	470

Berdasarkan Tabel 3, algoritma Random Forest menunjukkan kemampuan yang sangat baik dalam mengklasifikasikan nasabah tidak default dengan benar sebanyak 4.411 data, serta hanya menghasilkan 276 kesalahan klasifikasi (false positive). Namun demikian, algoritma ini masih menghasilkan 843 data default yang salah diklasifikasikan sebagai tidak default (false negative), sedangkan jumlah true positive tercatat sebanyak 470 data. Hal ini menunjukkan bahwa Random Forest sangat kuat dalam meminimalkan kesalahan pada kelas tidak default, tetapi relatif kurang sensitif dalam mendeteksi nasabah berisiko gagal bayar dibandingkan Naive Bayes.

d. Hasil Evaluasi Metrik

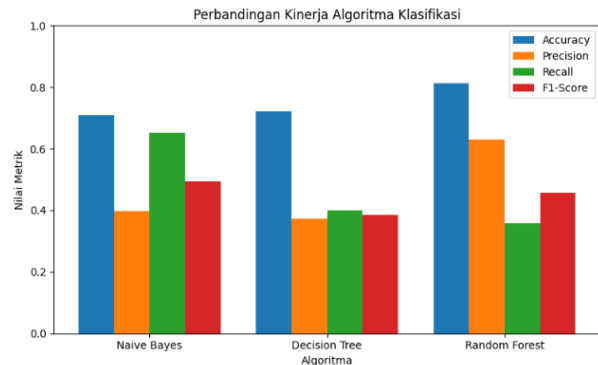
Setelah dilakukan pengujian menggunakan *testing*, selanjutnya dilakukan evaluasi kinerja model menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil evaluasi kinerja masing-masing algoritma disajikan pada Tabel 4.

Tabel 6. Hasil Perbandingan Algoritma

Model	Acc	Prec	Rec	F1
NB	0.708500	0.398416	0.651181	0.494363
DT	0.721833	0.373400	0.399848	0.386171
RF	0.813500	0.630027	0.357959	0.456532

Untuk mempermudah interpretasi hasil, perbandingan nilai *accuracy*, *precision*, *recall*, dan *F1-score* dari masing-masing algoritma divisualisasikan dalam bentuk

diagram batang sebagaimana ditunjukkan pada Gambar 3.



Gambar 3. Perbandingan Kinerja Algoritma Klasifikasi

Berdasarkan Gambar 3, algoritma Random Forest menghasilkan nilai *accuracy* tertinggi sebesar 0,813500 dan *precision* tertinggi sebesar 0,630027 dibandingkan algoritma lainnya. Sementara itu, algoritma Naive Bayes menghasilkan nilai *recall* tertinggi sebesar 0,651181 dan *F1-score* tertinggi sebesar 0,494363. Adapun algoritma Decision Tree menghasilkan nilai *accuracy* sebesar 0,721833, *precision* sebesar 0,373400, *recall* sebesar 0,399848, dan *F1-score* sebesar 0,386171.

Visualisasi tersebut menunjukkan adanya perbedaan karakteristik performa pada masing-masing algoritma. Random Forest cenderung unggul dalam ketepatan klasifikasi secara keseluruhan, sedangkan Naive Bayes memiliki kemampuan yang lebih baik dalam mendeteksi nasabah yang berpotensi mengalami gagal bayar (*default*).

Analisis Perbandingan Algoritma

Berdasarkan Tabel 4, algoritma Random Forest menghasilkan nilai accuracy tertinggi sebesar 0.813500, yang menunjukkan kemampuan terbaik dalam melakukan klasifikasi secara keseluruhan dibandingkan Naive Bayes dan Decision Tree. Selain itu, Random Forest juga memiliki nilai precision tertinggi sebesar 0.630027, yang mengindikasikan bahwa

model ini lebih akurat dalam memprediksi nasabah yang mengalami gagal bayar (*default*). Di sisi lain, algoritma Naive Bayes menunjukkan keunggulan pada metrik recall dengan nilai sebesar 0.651181, yang berarti algoritma ini lebih sensitif dalam mendeteksi nasabah yang benar-benar berisiko gagal bayar. Hal ini juga tercermin dari nilai F1-score tertinggi sebesar 0.494363 dibandingkan algoritma lainnya, yang menunjukkan keseimbangan yang lebih baik antara precision dan recall. Sementara itu, algoritma Decision Tree menghasilkan nilai accuracy sebesar 0.721833, precision sebesar 0.373400, recall sebesar 0.399848, dan F1-score sebesar 0.386171, yang secara keseluruhan lebih rendah dibandingkan dua algoritma lainnya.

Perbedaan kinerja tersebut menunjukkan bahwa Random Forest memiliki keunggulan dalam ketepatan klasifikasi secara keseluruhan karena merupakan metode ensemble yang mengombinasikan banyak pohon keputusan, sehingga mampu mengurangi variansi model dan menangkap pola kompleks pada data kredit. Sementara itu, Naive Bayes memiliki sensitivitas yang lebih baik terhadap kelas minoritas (*default*) karena pendekatan probabilitiknya, meskipun asumsi independensi antar fitur tidak sepenuhnya terpenuhi. Adapun Decision Tree memiliki keterbatasan karena model pohon tunggal cenderung sensitif terhadap variasi data dan ketidakseimbangan kelas.

Pembahasan dan Interpretasi

Hasil penelitian menunjukkan bahwa algoritma Random Forest menghasilkan performa terbaik berdasarkan nilai accuracy sebesar 0,813500 dan precision sebesar 0,630027. Temuan ini menunjukkan bahwa pendekatan ensemble learning yang digunakan oleh Random Forest mampu

meningkatkan stabilitas model dan menghasilkan klasifikasi yang lebih akurat dibandingkan algoritma Naive Bayes maupun Decision Tree. Kemampuan tersebut diperoleh karena Random Forest mengombinasikan banyak pohon keputusan sehingga dapat mengurangi variansi model dan menangkap pola yang lebih kompleks pada data risiko kredit.

Temuan penelitian ini sejalan dengan penelitian Rahmawati dan Handayani yang menunjukkan bahwa Random Forest mampu menghasilkan performa klasifikasi yang lebih tinggi pada kasus klasifikasi risiko kredit. Kesamaan hasil tersebut mengindikasikan bahwa metode ensemble learning memiliki kemampuan yang baik dalam menangani data kredit yang memiliki banyak atribut dan hubungan yang kompleks antar variabel.

Di sisi lain, algoritma Naive Bayes menghasilkan nilai recall tertinggi sebesar 0,651181 dan F1-score sebesar 0,494363. Hasil ini menunjukkan bahwa Naive Bayes memiliki kemampuan yang lebih baik dalam mendeteksi nasabah yang benar-benar berisiko mengalami gagal bayar (*default*). Temuan ini mendukung penelitian Ariadi yang menyatakan bahwa Naive Bayes efektif dalam mengenali pola probabilitik pada data kredit meskipun menggunakan asumsi independensi antar atribut. Sementara itu, algoritma Decision Tree menghasilkan performa yang lebih rendah dibandingkan kedua algoritma lainnya pada seluruh metrik evaluasi. Hal ini menunjukkan bahwa model pohon keputusan tunggal memiliki keterbatasan dalam menangani kompleksitas data risiko kredit dan lebih rentan terhadap variasi data maupun ketidakseimbangan kelas (*class imbalance*).

Hasil penelitian ini memiliki implikasi praktis bagi lembaga keuangan dalam pengelolaan risiko kredit nasabah kartu

ke kredit. Apabila tujuan utama lembaga keuangan adalah meminimalkan risiko kesalahan klasifikasi nasabah berisiko tinggi yang dinyatakan tidak berisiko (*false negative*), maka algoritma Naive Bayes dapat menjadi pilihan karena memiliki nilai recall yang lebih tinggi. Sebaliknya, apabila lembaga keuangan lebih mengutamakan ketepatan klasifikasi secara keseluruhan dan stabilitas model dalam mengelola data kredit berskala besar, maka algoritma Random Forest lebih direkomendasikan karena memiliki nilai accuracy dan precision yang lebih tinggi.

Model klasifikasi yang dihasilkan dalam penelitian ini berpotensi diintegrasikan ke dalam sistem *credit scoring* sebagai bagian dari sistem pendukung keputusan untuk membantu proses *credit risk assessment*. Dengan memanfaatkan hasil prediksi yang dihasilkan model, lembaga keuangan dapat melakukan penilaian risiko kredit secara lebih objektif, konsisten, dan berbasis data sehingga dapat meminimalkan potensi kerugian akibat gagal bayar nasabah.

Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, eksperimen hanya dilakukan menggunakan satu dataset sehingga hasil penelitian masih memiliki keterbatasan dalam generalisasi pada kondisi data yang berbeda. Kedua, penelitian ini belum menerapkan teknik penanganan ketidakseimbangan data seperti SMOTE, oversampling, maupun undersampling. Ketiga, model klasifikasi dibangun menggunakan parameter standar tanpa proses hyperparameter tuning. Keempat, penelitian ini belum melakukan pengujian signifikansi statistik terhadap perbedaan performa antar algoritma. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penggunaan beberapa dataset, optimasi parameter, teknik

penyeimbangan data, serta pengujian statistik untuk meningkatkan validitas hasil penelitian.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma Naive Bayes, Decision Tree, dan Random Forest berhasil diimplementasikan untuk melakukan prediksi risiko default nasabah kartu kredit. Hasil evaluasi menunjukkan bahwa Random Forest menunjukkan kinerja terbaik pada dataset dan skenario eksperimen yang digunakan dalam penelitian ini. Berdasarkan nilai accuracy dan precision tertinggi. Sementara itu, Naive Bayes menunjukkan kemampuan yang lebih baik dalam mendeteksi nasabah yang berpotensi mengalami gagal bayar yang ditunjukkan oleh nilai recall dan F1-score yang lebih tinggi.

Temuan ini menunjukkan bahwa pemilihan algoritma klasifikasi dalam *credit risk assessment* perlu mempertimbangkan tujuan analisis, baik untuk meningkatkan ketepatan klasifikasi secara keseluruhan maupun untuk meminimalkan risiko kesalahan dalam mendeteksi nasabah berisiko gagal bayar. Kontribusi penelitian ini terletak pada evaluasi komparatif tiga algoritma klasifikasi menggunakan pendekatan multi-metrik pada dataset kartu kredit berskala besar sehingga dapat menjadi referensi dalam pengembangan sistem pendukung keputusan kredit berbasis data.

Saran

Berdasarkan hasil penelitian yang diperoleh, penelitian selanjutnya disarankan untuk menerapkan teknik penanganan ketidakseimbangan data seperti SMOTE atau oversampling guna meningkatkan kemampuan model dalam mendeteksi

nasabah berisiko gagal bayar. Selain itu, optimasi hyperparameter dan pengujian signifikansi statistik dapat dilakukan untuk meningkatkan validitas hasil perbandingan antar algoritma. Penggunaan beberapa dataset dari institusi keuangan yang berbeda juga perlu dipertimbangkan agar hasil penelitian memiliki tingkat generalisasi yang lebih baik.

DAFTAR PUSTAKA

- Afandi, A., Bedi Agtriadi, H., Luqman, L., & Susanti, M. (2024). *Advanced Credit Scoring with Naive Bayes: Improving Accuracy and Reliability in Financial Risk Assessment*. *Jurnal E-Komtek (Elektro-Komputer-Teknik)*, 8(2), 399–409. <https://doi.org/10.37339/e-komtek.v8i2.2160>
- Anindita Puspa Ayu Prayogi, Altha Inas Shofyana, & Dewi Putriani. (2025). *Prediksi Credit Card Approval Menggunakan Algoritma Random Forest*. *Jurnal Ilmu Komputer Dan Multimedia*, 2(1), 21–25. <https://doi.org/10.46510/ilkomedia.v2i1.25>
- Antika, D. P., Fatekurohman, M., & Tirta, I. M. (2022). *Banking Credit Risk Analysis with Naive Bayes Approach and Cox Proportional Hazard*. *International Journal of Advanced Engineering Research and Science*, 9(8), 365–370. <https://doi.org/10.22161/ijaers.98.41>
- Ariadi, F. (2020). *Analisa Perbandingan Algoritma DT C.45 dan Naive Bayes Dalam Prediksi Penerimaan Kredit Motor Article History ABSTRAK*. *Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika (KERNEL)*, 1(1).
- Ayobami, A. O., Israel, O., Yusuf, J., & Eniola, A. M. (2023). *Predicting Credit Risk: Using Machine Learning Algorithms to Predict the Creditworthiness of Borrowers and Determine their Likelihood of Defaulting on Loans in Nigeria*. *International Journal of Recent Engineering Science*, 10(2), 46–53. <https://doi.org/10.14445/23497157/ijres-v10i2p107>
- Eka Yuni Titik Artaningsih, A. M. (2023). *Komparasi Algoritma Klasifikasi C4.5 Dan C4.5 Berbasis Particle Swarm Optimization Untuk Evaluasi Penentuan Kelayakan Pemberian Kredit Koperasi Di PT. Indah Kiat Pulp & Paper TBK*. *Scientia Sacra: Jurnal Sains, Teknologi Dan Masyarakat*, 3.
- Fatah, Z., & Khafifah, U. (2025). *Model Prediksi Risiko Kredit Nasabah Menggunakan Algoritma C4*. In *JAMASTIKA* (Vol. 4, Number 5). <https://doi.org/10.35473/jamastika.v4i2.4537>
- Gade, S. P., Nalkol, A., Ibitwar, S., & Kirpal, D. (2023). *Credit Risk Analysis Using Naive Bayes in Machine Learning*. In *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* (Vol. 11). <https://doi.org/doi.org/10.22214/ijraset.2023.52943>
- Hazizah, N., & Feranika, A. (2025). *Implementasi Algoritma Random Forest Dalam Klasifikasi Risiko Gagal Bayar Kartu Kredit Pada Nasabah Bank*. *Jurnal Manajemen Teknologi Dan Sistem Informasi (JMS)*, 5(1). <https://doi.org/10.33998/jms.v5i1>
- Kubra Bolukbas, E. T. (2025). *Machine Learning Applications in Credit Risk Prediction*. *Central Bank of the Republic of Türkiye*, 1–44.
- M. Rifqi Wirasena, J. W. (2024). *Penerapan Algoritma C4.5 Untuk Prediksi Kelayakan Pengajuan Kartu Kredit Visa Bagi Nasabah*. *TeknoIS: Jurnal Ilmiah Teknologi Informasi Dan Sains*, 14. <https://doi.org/http://doi.org/10.36350/jbs.v14i2>
- Novianti, T., Mandati, S. A., & Andana, E. K. (2023). *Peningkatan Evaluasi Risiko Kredit Menggunakan Decision Tree C*

4.5. Journal of Manufacturing in Industrial Engineering & Technology, 2(2), 1–9.
<https://doi.org/10.30651/mine-tech.v2i2.21749>

- Prayesy, P., Pujakesuma, A., & Qisthiano, M. R. (2025). *Evaluasi Kinerja Random Forest dan Naive Bayes untuk Prediksi Risiko Kredit Berdasarkan Pekerjaan Debitur*. *Explore*, 15(2), 114–120.
<https://doi.org/10.35200/ex.v15i2.175>
- Rahmawati, T. I., Handayani, I., & Yogyakarta, U. T. (2025). *Implementasi Random Forest Untuk Klasifikasi Kelayakan Kredit Nasabah Dalam Pengelolaan Risiko Keuangan Di Pt Bpr Bkk Wonogiri*. *Journal of Information Technology and Computer Science (INTECOMS)*, 8(6).
- Randy Fachrezi, M., Farhan, M., Aqila Aptanta, D., & Denis Wallidein, A. (2025). *Penerapan Algoritma Decision Tree Untuk Memprediksi Risiko Gagal Bayar Nasabah Kartu Kredit*. (14), 188–199.
<https://doi.org/10.26418/jaakfe.v14i2.101100>
- Rizky Saputra Cendikiawan, A. I. M. A. R. D. K. (2025). Analisis Sentimen terhadap Produk Zalora di Google Play Store Menggunakan Metode Random Forest. *Indonesian Journal of Innovation Studies*.