

Penerapan Algoritma Random Forest untuk Prediksi Kelayakan Kredit Pemilikan Rumah

Icha Sugiasari ¹, Nurjaya ²

¹Program Studi Teknik Informatika Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia
e-mail: ¹ichasugiasari12@gmail.com, ²dosen00370@unpam.ac.id

Info Artikel

Riwayat Artikel:

Received Jan 19, 2026
Revised April 21, 2026
Accepted May 24, 2026

Corresponding Author:

Icha Sugiasari
Email: ichasugiasari12@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract – Home Ownership Credit (KPR) is a banking financing product with a relatively high risk of default, therefore requiring an accurate and efficient creditworthiness assessment process. Conventional KPR evaluation procedures generally require a considerable amount of time due to the involvement of multiple assessment criteria. This study aims to implement the Random Forest algorithm in predicting the eligibility of Home Ownership Credit (KPR) applications to improve the efficiency and quality of credit decision-making. The data used in this study are secondary data obtained from the Loan Approval Prediction Dataset on the Kaggle platform, consisting of 4,269 credit application records. The research methods include data preprocessing, data exploration, encoding process, Random Forest model development, and performance evaluation using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC. The evaluation results show that the Random Forest model achieved an accuracy of 98.24%, precision of 98.13%, recall of 99.06%, F1-score of 98.59%, and an AUC value of 0.9983, indicating excellent classification performance. It can be concluded that the Random Forest algorithm is highly effective in predicting KPR eligibility and can support banks in accelerating the credit evaluation process while reducing the risk of non-performing loans.

Keywords: Home Ownership Credit (KPR); Random Forest; machine learning; credit eligibility prediction.

Abstrak Indonesia – Kredit Pemilikan Rumah (KPR) merupakan produk pembiayaan perbankan yang memiliki risiko gagal bayar cukup tinggi sehingga membutuhkan proses analisis kelayakan kredit yang akurat dan efisien. Proses penilaian secara konvensional umumnya memerlukan waktu yang relatif lama karena melibatkan banyak kriteria penilaian. Penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam memprediksi kelayakan pengajuan Kredit Pemilikan Rumah (KPR) guna meningkatkan efisiensi dan kualitas pengambilan keputusan kredit. Data yang digunakan merupakan data sekunder yang diperoleh dari Loan Approval Prediction Dataset pada platform Kaggle yang terdiri dari 4.269 data pengajuan kredit. Metode penelitian meliputi pengolahan data eksplorasi data, proses encoding, pembangunan model Random Forest, serta evaluasi kinerja model menggunakan metrik accuracy, precision, recall, F1-score, confusion matrix, dan ROC-AUC. Hasil evaluasi menunjukkan bahwa model Random Forest mencapai akurasi sebesar 98,24%, precision sebesar 98,13%, recall sebesar 99,06%, F1-score sebesar 98,59%, serta nilai AUC sebesar 0,9983. Kesimpulan penelitian ini menunjukkan bahwa algoritma Random Forest memiliki kinerja yang sangat baik dalam memprediksi kelayakan KPR dan berpotensi membantu perbankan dalam mempercepat proses evaluasi kredit serta mengurangi risiko kredit macet.

Kata Kunci: Kredit Pemilikan Rumah (KPR), Random Forest, machine learning, prediksi kelayakan kredit.

1. PENDAHULUAN

Kredit Pemilikan Rumah (KPR) merupakan fasilitas pembiayaan perbankan yang memungkinkan masyarakat membeli rumah melalui sistem angsuran dengan rumah sebagai jaminan hingga kredit lunas [1]. KPR menjadi salah satu produk pembiayaan yang banyak digunakan masyarakat karena memberikan kemudahan dalam memiliki hunian tanpa harus menyediakan dana tunai dalam jumlah besar di awal. Selain itu, tenor pembiayaan yang panjang menjadikan KPR sebagai solusi kepemilikan rumah yang banyak diminati [2].

Namun, pemberian KPR memiliki risiko kredit macet yang dapat merugikan pihak perbankan apabila proses penilaian kelayakan kredit tidak dilakukan secara tepat. Penilaian kelayakan kredit secara konvensional umumnya membutuhkan waktu yang relatif lama karena melibatkan berbagai atribut dan kriteria penilaian calon nasabah [3]. Oleh karena itu, diperlukan metode yang mampu membantu proses pengambilan keputusan kredit secara lebih cepat, objektif, dan akurat.

Perkembangan teknologi machine learning memberikan peluang dalam membangun sistem prediksi berbasis data untuk membantu proses klasifikasi kelayakan kredit [7]-[8]. Salah satu algoritma yang banyak digunakan dalam proses klasifikasi adalah Random Forest. Algoritma ini merupakan metode ensemble learning yang menggabungkan sejumlah decision tree untuk menghasilkan model prediksi yang lebih stabil dan memiliki kemampuan generalisasi yang baik [9]-[10].

Beberapa penelitian sebelumnya telah menerapkan algoritma machine learning pada kasus prediksi kelayakan kredit. Penelitian Joshua Valens Andrian dan Charitas Fibriani menggunakan algoritma Naïve Bayes dan memperoleh akurasi sebesar 86% [4]. Penelitian Ainun Mardiah Hasibuan dan Sriani menggunakan algoritma K-Nearest Neighbor dengan akurasi sebesar 90% [5]. Selain itu, penelitian Noval Firmansah, Uce Indahyanti, dan Ade Eviyanti menggunakan Random Forest dan memperoleh akurasi sebesar 93,28% [6].

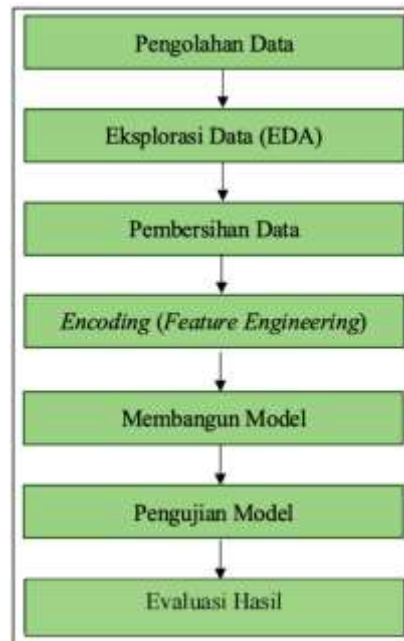
Berdasarkan penelitian terdahulu tersebut, penelitian ini menerapkan algoritma Random Forest untuk memprediksi kelayakan Kredit Pemilikan Rumah (KPR). Penelitian difokuskan pada pembangunan model machine learning menggunakan dataset pengajuan kredit guna membantu mempercepat proses evaluasi kredit dan mengurangi risiko kredit macet.

2. METODE PENELITIAN

Penelitian ini menggunakan dataset yang merupakan data sekunder yang diperoleh dari platform Kaggle dengan judul “Loan Approval Prediction Dataset”. Dataset terdiri dari 4.269 record dengan 12 atribut sebagai fitur input dan 1 atribut output yaitu loan_status. Seluruh atribut finansial pada dataset menggunakan satuan mata uang Rupee India (INR).

Tahapan Penelitian

Penelitian ini menggunakan beberapa tahapan pengolahan data, yaitu pengumpulan data, pembersihan data, eksplorasi data, encoding data kategorikal, pembangunan model Random Forest, dan evaluasi model. Setiap tahapan penelitian dilakukan secara berurutan untuk memastikan data yang digunakan memiliki kualitas yang baik sebelum masuk ke proses pemodelan. Tahap preprocessing dilakukan untuk mengurangi kesalahan data dan memastikan seluruh atribut dapat diproses oleh algoritma Random Forest. Setelah proses pengolahan selesai, model kemudian dilatih menggunakan data latih dan dilakukan evaluasi menggunakan data evaluasi untuk mengetahui kemampuan model dalam melakukan klasifikasi kelayakan kredit secara akurat.



Gambar 1. Rancangan Penelitian

a. Pengolahan Data

Tahap awal dilakukan dengan menyiapkan dataset Loan Approval Prediction Dataset dari Kaggle sebelum proses analisis dan pemodelan. Proses yang dilakukan meliputi merapikan nama kolom agar konsisten serta menghapus kolom `loan_id` karena tidak berpengaruh terhadap prediksi. Proses pengolahan data dilakukan untuk meningkatkan kualitas dataset sebelum digunakan pada tahap pelatihan model. Data yang telah dirapikan membantu algoritma dalam mengenali pola hubungan antar atribut sehingga proses klasifikasi dapat dilakukan dengan lebih optimal.

b. Eksplorasi Data (EDA)

Eksplorasi data dilakukan untuk memahami karakteristik dataset dengan melihat jumlah data, tipe data, distribusi atribut, dan hubungan awal antar variabel terhadap `loan_status`. Tahap ini membantu peneliti mengenali pola data sebelum proses pemodelan. Melalui proses eksplorasi data, peneliti dapat mengetahui kondisi dataset serta memahami karakteristik masing-masing atribut yang berpengaruh terhadap status kelayakan kredit. Tahap ini juga membantu dalam menentukan proses pengolahan data yang sesuai sebelum dilakukan pemodelan.

c. Pembersihan Data

Pembersihan data dilakukan melalui pengecekan data kosong dan duplikat. Hasilnya menunjukkan dataset tidak memiliki nilai kosong maupun data duplikat. Outlier pada atribut pendapatan dan aset tetap dipertahankan agar kondisi finansial tetap terwakili. Tidak ditemukannya missing value maupun data duplikat menunjukkan bahwa dataset memiliki kualitas data yang cukup baik untuk digunakan dalam proses pelatihan model machine learning.

d. Encoding (Feature Engineering)

Atribut kategorikal seperti `education`, `self_employed`, dan `loan_status` diubah menjadi bentuk numerik agar dapat diproses oleh algoritma Random Forest.

e. Membangun Model

Model dibangun menggunakan algoritma Random Forest Classifier dengan memanfaatkan data latih yang telah dipersiapkan. Algoritma bekerja melalui pembentukan banyak pohon keputusan dan menentukan hasil prediksi berdasarkan suara terbanyak.

f. Pengujian Model

Pengujian awal dilakukan menggunakan data validasi internal untuk mengetahui performa model sebelum evaluasi akhir dilakukan.

g. Evaluasi Hasil

Evaluasi model dilakukan menggunakan data evaluasi (uji akhir) dengan metrik accuracy, precision, recall, dan F1-score. Selain itu digunakan Confusion Matrix dan ROC-AUC untuk menilai kemampuan model dalam membedakan pengajuan kredit layak dan tidak layak. Dataset terdiri dari 3415 data latih dan 854 data evaluasi dengan total 11 kolom.

Table I. Dimensi Dataset

Dataset	Jumlah
Data Latih	3415 Baris data & 11 Kolom
Data Evaluasi	854 Baris data & 11 Kolom

Dataset dibagi menjadi data latih dan data evaluasi, di mana data latih mencakup data pelatihan dan validasi internal model, sedangkan data evaluasi digunakan sebagai data uji akhir untuk menilai kinerja dan kemampuan generalisasi model.

Perangkat Penelitian

Perangkat keras yang digunakan memiliki spesifikasi minimal Intel Core i5 generasi ke-10 atau setara, RAM minimal 8 GB, dan penyimpanan SSD minimal 256 GB. Perangkat lunak yang digunakan meliputi Python 3.9, Jupyter Notebook, serta beberapa library Python seperti Pandas, NumPy, Scikit-Learn, Matplotlib, dan Seaborn.

Random Forest

Random Forest merupakan algoritma ensemble learning yang menggabungkan sejumlah decision tree untuk menghasilkan model klasifikasi yang lebih akurat dan stabil. Algoritma ini membangun banyak pohon keputusan menggunakan teknik bootstrap sampling dan random feature selection. Prediksi akhir diperoleh menggunakan mekanisme majority voting berdasarkan hasil klasifikasi dari seluruh pohon keputusan.

Pada implementasi penelitian ini, model Random Forest dibangun menggunakan beberapa parameter utama, yaitu `n_estimators` sebesar 200, `criterion` menggunakan `gini`, `bootstrap` bernilai `true`, dan `max_features` menggunakan `sqrt`. Seperti pada tabel berikut:

Table II. Parameter Model Random Forest

Parameter	Nilai
<code>n_estimators</code>	200
<code>criterion</code>	<code>gini</code>
<code>class_weight</code>	<code>balanced</code>
<code>random_state</code>	42
<code>bootstrap</code>	<code>True</code>
<code>max_depth</code>	<code>Default (None)</code>
<code>min_samples_split</code>	2
<code>min_samples_leaf</code>	1
<code>max_features</code>	<code>sqrt</code>

Penggunaan parameter tersebut bertujuan untuk menghasilkan model yang stabil dan mampu melakukan klasifikasi secara optimal. Penambahan jumlah pohon keputusan pada Random Forest membantu meningkatkan kemampuan generalisasi model serta mengurangi risiko overfitting terhadap data pelatihan.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset Loan Approval Prediction Dataset yang terdiri dari 4.269 data pengajuan kredit dengan atribut finansial dan identitas pemohon kredit. Pada tahap awal dilakukan proses preprocessing data yang meliputi pengecekan missing value, penghapusan atribut yang tidak relevan, serta encoding terhadap atribut kategorikal agar dapat diproses menggunakan algoritma machine learning.

Setelah proses preprocessing selesai, data dibagi menjadi data latih dan data evaluasi dengan perbandingan 80:20. Data latih digunakan untuk proses pelatihan model Random Forest, sedangkan data evaluasi digunakan untuk menguji performa model terhadap data baru.

Model Random Forest dibangun menggunakan parameter `n_estimators` sebesar 200 dengan criterion `gini` dan `max_features` menggunakan `sqrt`. Parameter tersebut digunakan untuk menghasilkan model klasifikasi yang stabil dan mampu mengurangi overfitting.

Hasil Pengujian Data Validasi

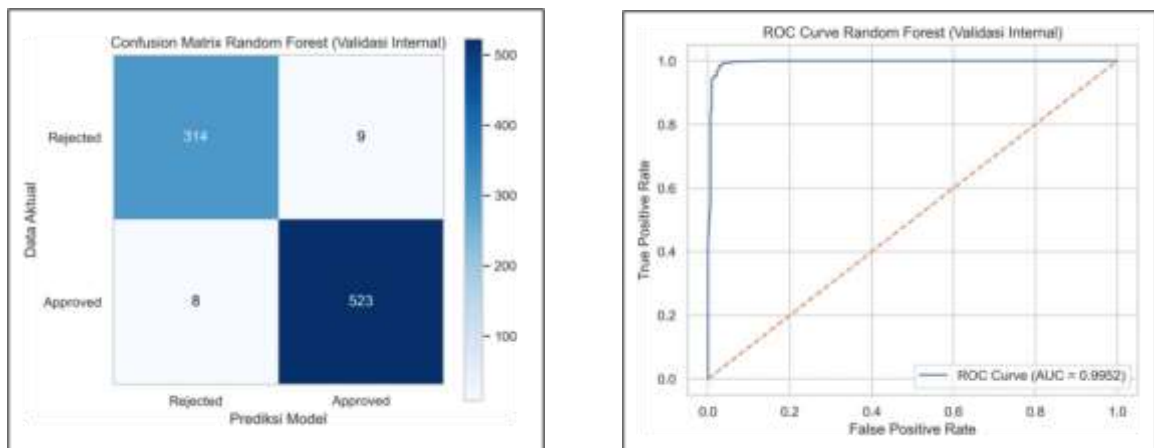
Hasil pengujian pada data validasi menunjukkan performa model yang sangat baik. Model memperoleh accuracy sebesar 98,01%, precision sebesar 98,31%, recall sebesar 98,49%, dan F1-score sebesar 98,40%.

Table III. Hasil Pengujian Data Validasi

Metrik	Nilai
Accuracy	0.9801
Precision	0.9831
Recall	0.9849
F1-Score	0.9840

Nilai accuracy sebesar 98,01% menunjukkan bahwa sebagian besar data validasi berhasil diklasifikasikan dengan benar oleh model Random Forest. Nilai precision sebesar 98,31% menunjukkan bahwa prediksi terhadap kelas Approved memiliki tingkat ketepatan yang sangat tinggi. Selain itu, recall sebesar 98,49% menunjukkan bahwa model mampu mengenali hampir seluruh data yang termasuk ke dalam kelas Approved. Nilai F1-score sebesar 98,40% menunjukkan keseimbangan yang baik antara precision dan recall dalam proses klasifikasi.

Selain menggunakan metrik evaluasi, pengujian juga dilakukan menggunakan confusion matrix dan ROC Curve. Berdasarkan confusion matrix, sebagian besar data berhasil diklasifikasikan dengan benar ke dalam kelas Approved maupun Rejected. Hasil ROC Curve pada data validasi menghasilkan nilai AUC sebesar 0,9952 yang menunjukkan kemampuan klasifikasi yang sangat baik. Seperti pada gambar berikut:



Gambar 2. Confusion Matrix dan ROC Curve Data Validasi

Berdasarkan confusion matrix pada data validasi, sebagian besar data berhasil diprediksi sesuai dengan kelas aslinya. Kurva ROC yang mendekati sudut kiri atas juga menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dalam membedakan kelas Approved dan Rejected.

Hasil Pengujian Data Evaluasi

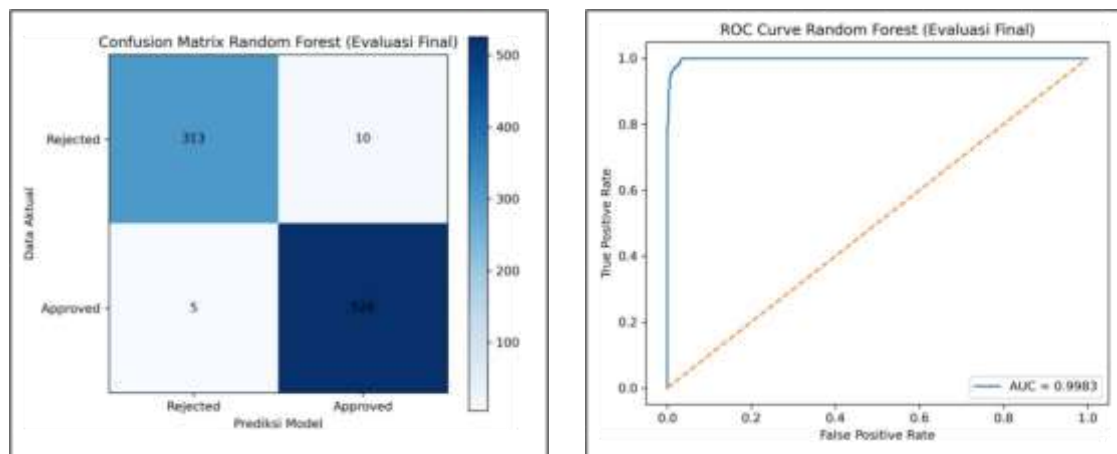
Pengujian akhir dilakukan menggunakan data evaluasi independen sebanyak 854 data. Berdasarkan hasil pengujian, model Random Forest memperoleh accuracy sebesar 98,24%, precision sebesar 98,13%, recall sebesar 99,06%, dan F1-score sebesar 98,59%.

Table IV. Hasil Pengujian Data Evaluasi

Metrik	Nilai
Accuracy	0.9824
Precision	0.9813
Recall	0.9906
F1-Score	0.9859

Hasil pengujian pada data evaluasi menunjukkan bahwa model Random Forest tetap mampu mempertahankan performa klasifikasi yang tinggi terhadap data baru. Tingginya nilai recall menunjukkan bahwa model mampu mengenali sebagian besar pengajuan kredit yang layak dengan baik sehingga potensi kesalahan klasifikasi dapat diminimalkan.

Berdasarkan confusion matrix pada data evaluasi, sebanyak 526 data Approved berhasil diprediksi dengan benar sebagai Approved dan 313 data Rejected berhasil diprediksi dengan benar sebagai Rejected. Selain itu, terdapat 10 data Rejected yang diprediksi sebagai Approved dan 5 data Approved diprediksi sebagai Rejected.



Gambar 3. Confusion Matrix dan ROC Curve Data Evaluasi

Nilai ROC-AUC sebesar 0,9983 menunjukkan bahwa model memiliki tingkat diskriminasi yang sangat baik. Semakin mendekati nilai 1, maka semakin baik kemampuan model dalam membedakan antara kelas Approved dan Rejected.

Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu memberikan performa klasifikasi yang sangat baik dalam memprediksi kelayakan Kredit Pemilikan Rumah (KPR). Tingginya nilai accuracy menunjukkan bahwa model mampu melakukan prediksi dengan tingkat kesalahan yang rendah terhadap data pengajuan kredit.

Nilai precision yang tinggi menunjukkan bahwa sebagian besar data yang diprediksi sebagai Approved benar-benar termasuk ke dalam kelas Approved. Hal ini penting dalam proses penilaian kredit karena dapat membantu mengurangi risiko pemberian kredit kepada nasabah yang tidak layak. Nilai recall sebesar 99,06% menunjukkan bahwa model mampu mengenali hampir seluruh data Approved dengan baik. Tingginya recall menunjukkan bahwa model memiliki kemampuan yang baik dalam mengurangi kesalahan klasifikasi terhadap calon nasabah yang sebenarnya layak menerima kredit. Nilai F1-score sebesar 98,59% menunjukkan keseimbangan yang baik antara precision dan recall. Hasil tersebut menandakan bahwa model tidak hanya memiliki ketepatan prediksi yang tinggi tetapi juga mampu mengenali data positif secara konsisten. Hasil evaluasi ROC-AUC sebesar 0,9983 menunjukkan bahwa model Random Forest memiliki kemampuan diskriminasi yang sangat baik dalam membedakan kelas Approved dan Rejected. Semakin mendekati nilai 1, maka semakin baik kemampuan model dalam melakukan klasifikasi.

Penggunaan algoritma Random Forest pada penelitian ini juga menunjukkan bahwa metode ensemble learning mampu meningkatkan stabilitas model dibandingkan penggunaan satu decision tree tunggal. Kombinasi beberapa pohon keputusan menghasilkan prediksi yang lebih konsisten serta mampu mengurangi risiko overfitting.

Hasil penelitian ini menunjukkan bahwa algoritma Random Forest mampu memanfaatkan data historis pengajuan kredit untuk membangun pola klasifikasi yang baik. Penggunaan banyak pohon keputusan memungkinkan model menghasilkan prediksi yang lebih stabil dibandingkan penggunaan satu decision tree tunggal. Selain itu, penerapan Random Forest pada penelitian ini menunjukkan bahwa metode machine learning dapat membantu proses evaluasi kredit menjadi lebih efisien. Model dapat digunakan sebagai sistem pendukung keputusan untuk membantu pihak perbankan dalam melakukan analisis kelayakan kredit secara lebih cepat dan objektif. Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma Random Forest dapat digunakan sebagai solusi dalam membantu proses evaluasi kelayakan Kredit Pemilikan Rumah (KPR). Model yang dihasilkan mampu membantu proses pengambilan keputusan kredit menjadi lebih cepat, objektif, dan berbasis data sehingga berpotensi membantu perbankan dalam mengurangi risiko kredit macet.

4. KESIMPULAN

Berdasarkan hasil implementasi dan pengujian, algoritma Random Forest berhasil diterapkan dalam menentukan kelayakan pengajuan Kredit Pemilikan Rumah (KPR) melalui tahapan pengolahan data, pemodelan, dan evaluasi menggunakan data historis pemohon kredit. Model yang dibangun mampu mempercepat proses pengambilan keputusan kelayakan kredit dengan performa klasifikasi yang sangat baik. Hasil evaluasi menunjukkan bahwa model Random Forest memperoleh accuracy sebesar 98,24%, precision sebesar 98,13%, recall sebesar 99,06%, F1-score sebesar 98,59%, dan ROC-AUC sebesar 0,9983. Hasil tersebut menunjukkan bahwa algoritma Random Forest memiliki kemampuan yang sangat baik dalam mengklasifikasikan pengajuan kredit secara akurat dan stabil sehingga berpotensi membantu perbankan dalam mengurangi risiko kredit macet.

DAFTAR PUSTAKA

- [1] A. Mujib, “Analisis Restrukturisasi Pembiayaan KPR Pada Masa Pandemi Covid-19,” *Jurnal Masharif Al-Syariah: Jurnal Ekonomi dan Perbankan Syariah*, vol. 8, no. 3, pp. 716–723, 2023
- [2] M. Ridwan, F. E. Rahmatunnisa, S. Azzahra, T. Listiani, dan T. Agustine, “Analisis Transaksi Kredit Pemilikan Rumah (KPR) Dalam Tinjauan Hukum Islam Di Bank Tabungan Negara Syariah Cirebon,” *Ecopreneur: Jurnal Ekonomi dan Bisnis Islam*, vol. 2, no. 1, pp. 152–162, 2021. <http://journal.bungabangsacirebon.ac.id/index.php/ecopreneur>
- [3] A. Wijaya dan D. P. Sari, “Aspek-aspek penilaian dalam pemberian kredit bank,” *Jurnal Lex Privatum*, vol. 3, no. 4, pp. 1–11, 2015. <https://ejournal.unsrat.ac.id/index.php/lexprivatum/article/view/8986>
- [4] J. V. Andrian dan C. Fibriani, “Model Klasifikasi Kelayakan Pengajuan Kredit Bank Menggunakan Naïve Bayes,” *Jurnal Sifo Mikroskil*, vol. 26, no. 2, pp. 275–288, 2025, doi: 10.55601/jsm.v26i2.1803
- [5] A. M. Hasibuan dan Sriani, “Klasifikasi kelayakan peminjaman nasabah menggunakan algoritma K-Nearest Neighbor (K-NN),” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1525–1532, 2024, doi: 10.57152/malcom.v4i4.1682.
- [6] N. Firmansah, U. Indahyanti, dan A. Eviyanti, “Prediksi kelayakan pemberian kredit menggunakan metode random forest,” *Jurnal Ilmiah KOMPUTASI*, vol. 22, no. 4, 2023, doi: 10.32409/jikstik.22.4.3515.
- [7] M. N. Fahmi, “Implementasi Machine Learning Menggunakan Python Library: Scikit-Learn (Supervised dan Unsupervised Learning),” *Sains Data: Jurnal Studi Matematika dan Teknologi*, vol. 1, no. 2, pp. 87–96, 2023. doi: 10.52620/sainsdata.v1i2.9.
- [8] I. Anas, A. P. Widodo, and A. Setiawan, “Systematic Literature Review: Peran Machine Learning dalam Manajemen Sumber Daya Manusia,” *OPTIMAL Jurnal Ekonomi dan Manajemen*, vol. 5, no. 4, pp. 650–665, 2025, doi: 10.55606/optimal.v5i4.8089.
- [9] O. Y. Inonu, K. Magda, dan Amarudin, “Analisis kinerja algoritma Random Forest dengan model machine learning pada dataset penyakit diabetes,” *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, vol. 15, no. 1, hlm. 1–7, 2025, doi: 10.36448/expert.v15i1.4312.
- [10] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Sebastopol, CA, USA: O’Reilly Media, 2019