*Journal Software, Hardware and Information Technology*
*p-ISSN: 2808-3385 | e-ISSN: 2776-8961*
*Volume 5 Issue 2, June 2025*

# Perbandingan Kinerja Model Pembelajaran Mesin Random Forest dan K-Nearest Neighbor (KNN) untuk Prediksi Risiko Kredit pada Layanan Pinjaman Online

*Performance Comparison of Random Forest and K-Nearest Neighbor (KNN) Machine Learning Models for Credit Risk Prediction in Online Lending Services*

Santi Prayudani, Yous Sibarani, Azrizal Salam, Arif Ridho Lubis
Jurusan Teknik Komputer dan Informatika, Politeknik Negeri Medan, Medan, Indonesia
santiprayudani@polmed.ac.id, yousmarkuss.sibarani@students.polmed.ac.id, azrizalsalam@students.polmed.ac.id, arifridho@polmed.ac.id

*Notes*

*Submitted: 14-06-2025*
*Revised: 21-06-2025*
*Accepted: 23-06-2025*
*Published: 30-06-2025*

**How to cite:** S. Prayudani, Y. Sibarani, A. Salam, and A. R. Lubis, "Perbandingan Kinerja Model Pembelajaran Mesin Random Forest dan K-Nearest Neighbor (KNN) untuk Prediksi Risiko Kredit pada Layanan Pinjaman Online," *Journal Software, Hardware and Information Technology (SHIFT)*, vol. 5, no. 2, pp. 123–132, Jun. 2025, doi: 10.24252/shift.v5i2.204.

**DOI**: 10.24252/shift.v5i2.204

Copyright 2025© the Author(s)

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

## ABSTRACT

This study aims to compare the performance of two popular machine learning algorithms, Random Forest and K-Nearest Neighbor (KNN), in predicting creditworthiness in online lending systems. The research uses the publicly available Loan Approval Prediction Dataset from Kaggle, which contains borrower profiles such as employment status, number of dependents, annual income, loan amount, loan term, and credit score. Data preprocessing included cleaning, handling missing values, outlier removal, and transformation through normalization and encoding. The dataset was divided into 80% training data and 20% testing data. Random Forest was configured with 100 decision trees and unlimited depth, while KNN used an optimal k value of 5 determined by grid search. Model performance was evaluated using accuracy, precision, recall, and F1-score. The results showed that Random Forest outperformed KNN with consistently higher values (97%) across all metrics, demonstrating strong stability and superior pattern recognition capabilities. KNN, with an accuracy of 89%, still showed good performance and can be considered a lightweight alternative for simpler applications.

**Keywords:** classification, credit risk prediction, k-nearest neighbor, machine learning, random forest.

## ABSTRAK

Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma pembelajaran mesin yang populer, yaitu Random Forest dan K-Nearest Neighbor (KNN), dalam memprediksi kelayakan kredit pada sistem pinjaman online. Data yang digunakan adalah *Loan Approval Prediction Dataset* yang tersedia di platform Kaggle, yang berisi profil peminjam seperti status pekerjaan, jumlah tanggungan, pendapatan tahunan, jumlah pinjaman, jangka waktu pinjaman, dan skor

kredit. Tahap pra-pemrosesan data meliputi pembersihan, penanganan nilai kosong, penghapusan outlier, serta normalisasi dan *encoding*. Dataset dibagi menjadi 80% data latih dan 20% data uji. Model Random Forest dikonfigurasi dengan 100 pohon keputusan dan kedalaman tak terbatas, sedangkan model KNN menggunakan nilai *k* optimal sebesar 5 yang diperoleh melalui uji coba *grid search*. Evaluasi kinerja model dilakukan dengan empat metrik utama: akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Random Forest memiliki performa lebih unggul dengan nilai yang konsisten tinggi (97%) pada semua metrik, menandakan kestabilan dan kemampuan mengenali pola yang kuat. Meskipun demikian, KNN tetap menunjukkan performa baik dengan akurasi 89% dan dapat dijadikan alternatif yang lebih ringan untuk aplikasi yang sederhana.

**Kata kunci:** k-nearest neighbor, klasifikasi, pembelajaran mesin, prediksi risiko kredit, random forest.

## 1. PENDAHULUAN

Perkembangan teknologi digital yang pesat telah membawa perubahan besar dalam sektor keuangan, salah satunya adalah kemunculan sistem pinjaman online atau yang lebih dikenal sebagai pinjol. Perkembangan ini telah menggeser model bisnis keuangan dari sistem konvensional menuju sistem digital yang memungkinkan pengguna pinjol (pinjaman online) melakukan transaksi keuangan tanpa harus bertatap muka. Proses seperti peminjaman uang yang dahulu memerlukan kehadiran fisik kini dapat dilakukan hanya dalam hitungan detik melalui perangkat digital [1]. Perkembangan ini memberikan kemudahan, namun juga menimbulkan tantangan baru, khususnya terkait dengan risiko dan literasi masyarakat terhadap layanan keuangan digital seperti pinjaman online. Seiring dengan hal itu, pinjaman online menunjukkan bahwa layanan pinjaman digital mampu memperluas jangkauan pembiayaan mikro dengan memanfaatkan integrasi teknologi informasi secara efisien dan adaptif untuk karakteristik usaha kecil dan kondisi ekonomi peminjam [2].

Meskipun sistem penilaian statistik telah lama digunakan untuk mengevaluasi risiko kredit, pendekatan ini dinilai semakin terbatas dalam menangkap kompleksitas perilaku finansial modern. Model-model ini mengandalkan variabel tetap seperti riwayat pembayaran dan rasio utang terhadap pendapatan, yang sifatnya statis dan tidak responsif terhadap perubahan ekonomi yang cepat berubah [3]. Akibatnya, banyak pengguna yang sebenarnya layak tetapi tidak dapat melakukan peminjaman karena adanya celah dalam sistem keuangan terutama bagi yang belum memiliki riwayat kredit sebelumnya. Dalam industri keuangan, pengelolaan risiko kredit yang efektif sangat penting untuk menjaga stabilitas sistem perbankan dan menghindari akumulasi pinjaman yang bermasalah [4].

Teknologi pembelajaran mesin telah terbukti mampu meningkatkan kecepatan dan akurasi dalam penilaian kelayakan kredit, terutama dalam menjangkau individu yang belum memiliki riwayat kredit [5]. Penggunaan pembelajaran mesin memungkinkan premrosesan data peminjam yang lebih dinamis dan variatif, sehingga mampu menangkap indikator risiko yang tidak terdeteksi oleh pendekatan statistik konvensional. Pemanfaatan kecerdasan buatan memungkinkan institusi keuangan untuk memanfaatkan data tak terstruktur, seperti jejak

digital, aktivitas media sosial, hingga pola transaksi untuk mengidentifikasi risiko kredit secara presisi [6]. Hal ini sejalan dengan temuan oleh peneliti terdahulu, yang menekankan pentingnya pemanfaatan jejak digital dan pendekatan machine learning dalam segmentasi risiko kredit yang lebih akurat [7].

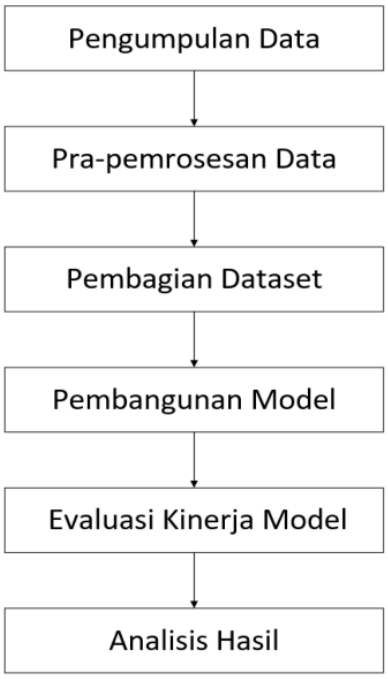
Di era digital, keberadaan data menjadi peluang besar untuk meningkatkan kualitas penilaian risiko kredit. Teknologi pembelajaran mesin menawarkan solusi yang dapat membantu untuk menganalisis kelayakan kredit secara lebih akurat dan efisien [8]. Algoritma pembelajaran mesin mampu mengekstraksi pola tersembunyi dari data yang besar dan tidak terstruktur. Dengan kelebihan ini memungkinkan untuk melatih model pembelajaran mesin yang dapat mengukur risiko kredit yang lebih akurat, efisien, dan responsif terhadap perubahan ekonomi peminjam. Model pembelajaran mesin seperti Random Forest telah menunjukkan performa prediktif yang lebih tinggi dibandingkan model tunggal [9].

Terdapat berbagai algoritma yang telah digunakan oleh para peneliti untuk melakukan prediksi, dua di antaranya adalah Random Forest dan K-Nearest Neighbor (KNN). Random Forest memiliki keunggulan untuk menangani dataset yang kompleks dan memiliki kinerja yang baik dalam tugas klasifikasi. Di sisi lain KNN adalah algoritma yang sederhana dan mudah diimplementasikan. Metode klasifikasi seperti Random Forest dan KNN merupakan alat yang umum digunakan dalam prediksi risiko kredit karena mampu memetakan kemungkinan kelas risiko dari peminjam secara akurat dan efisien berdasarkan pola historis data kredit [10].

Penelitian ini bertujuan untuk melakukan perbandingan kinerja antara algoritma Random Forest dan KNN dalam memprediksi kelayakan kredit pada sistem pinjaman online. Kedua algoritma ini dipilih karena memiliki keunggulan masing-masing, Random Forest dikenal mampu menangani data yang kompleks dan heterogen dengan stabilitas yang tinggi, sementara KNN dikenal karena kesederhanaannya dan kecepatan komputasinya pada dataset yang lebih kecil dan homogen. Evaluasi kinerja dilakukan secara komprehensif menggunakan empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score, yang secara bersama-sama memberikan gambaran yang lebih lengkap mengenai efektivitas masing-masing model. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan sistem pendukung keputusan di sektor keuangan, khususnya dalam mengoptimalkan proses analisis risiko kredit yang berbasis data dan pembelajaran mesin. Penelitian ini diharapkan mampu menjadi referensi praktis bagi penyedia layanan pinjaman online dalam meningkatkan akurasi penilaian risiko peminjam dan memperkuat kualitas layanan keuangan digital yang ditawarkan.

## 2. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan yang terstruktur dan sistematis guna mencapai tujuan utama, yaitu membandingkan kinerja algoritma Random Forest dan K-Nearest Neighbor (KNN) dalam memprediksi kelayakan kredit pada sistem pinjaman online. Secara garis besar metode penelitian ini seperti yang ada pada Gambar 1 di bawah ini:




**Gambar 1.** Metode Penelitian

### 2.1. Pengumpulan Data

Tahap pertama adalah pengumpulan data. Penelitian ini menggunakan data *Loan Approval Prediction Dataset* sebagai dataset yang diperoleh melalui platform Kaggle. Dataset ini berisi data historis pengguna pinjaman online, seperti status pekerjaan, jumlah tanggungan, pendapatan, aset yang dimiliki, dan data demografis lainnya. Pemilihan dataset merupakan langkah penting dalam penelitian klasifikasi kredit, karena kualitas dan variasi data dapat memengaruhi keakuratan model secara signifikan [11].

### 2.2. Pra-pemrosesan Data

Tahap kedua adalah pra-pemrosesan data. Meskipun dataset telah bersih dari segi struktural, dilakukan pengecekan ulang secara teliti untuk memastikan tidak terdapat data kosong, atau data duplikat yang dapat mengganggu proses pembelajaran model. Tahapan ini juga mencakup normalisasi data numerik agar setiap fitur berada dalam skala yang sama, yang akan memudahkan algoritma dalam mengenali pola-pola tersembunyi dan meningkatkan stabilitas prediksi. Selain itu, data kategorikal diubah menjadi format numerik melalui proses *encoding*, yang bertujuan untuk memastikan semua variabel dapat dimanfaatkan secara optimal dalam proses pelatihan model.

### 2.3. Pembagian Dataset

Tahap ketiga adalah pembagian dataset. Data yang telah melalui tahap pra-pemrosesan kemudian dibagi menjadi dua bagian utama yaitu data latih sebesar 80% dan data uji sebesar 20% menggunakan metode *train-test split* dengan parameter *random_state=42* dan stratify untuk menjaga proporsi kelas seimbang antara data latih dan data uji. Tujuan utama dari pembagian ini adalah untuk mengevaluasi performa model secara lebih akurat dan objektif. Model akan dilatih menggunakan data latih agar dapat mempelajari pola dan hubungan antarvariabel, lalu diuji pada data uji yang belum pernah dilihat oleh model sebelumnya.

### 2.4. Pembangunan Model

Tahap keempat adalah pembangunan model. Pada tahap ini, dua algoritma utama diterapkan, yaitu Random Forest dan K-Nearest Neighbor (KNN), untuk melakukan klasifikasi kelayakan kredit. Model Random Forest dibangun menggunakan parameter dasar, dengan jumlah *n_estimators* sebanyak 100 pohon keputusan. Parameter *max_depth* disetel tak terbatas agar setiap pohon dapat mempelajari pola yang kompleks dan beragam dalam data.

Untuk model KNN, nilai *k* yang digunakan ditentukan melalui uji coba sistematis menggunakan teknik *grid search* dengan rentang nilai k dari 3 hingga 15. Teknik ini membantu untuk menemukan nilai *k* yang paling optimal untuk menghasilkan akurasi tertinggi pada data latih. Penerapan serupa dilakukan oleh peneliti terdahulu, yang mengimplementasikan algoritma KNN dalam penilaian risiko kredit koperasi dan memperoleh akurasi sebesar 85% [12]. Model KNN menggunakan fungsi jarak *Euclidean* sebagai pengukur kedekatan antar data, yang menjadi dasar dalam menentukan klasifikasi risiko kredit. Kedua model ini kemudian dilatih menggunakan data latih yang telah diproses, sehingga dapat menangkap pola dan hubungan antar variabel yang relevan.

### 2.5. Evaluasi Kinerja Model

Tahap kelima adalah evaluasi kinerja model. Model Random Forest dan KNN yang telah dilatih kemudian dievaluasi menggunakan data uji yang tidak pernah digunakan selama pelatihan, untuk memastikan bahwa hasil evaluasi mencerminkan kemampuan generalisasi model secara objektif. Evaluasi model dilakukan dengan metode hold-out dengan menggunakan data uji independent dengan total 20% data yang ada pada dataset. Evaluasi dilakukan dengan empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score. Akurasi digunakan untuk mengukur persentase prediksi yang benar dari keseluruhan data uji. Presisi menunjukkan seberapa relevan hasil prediksi positif yang dibuat oleh model. Recall menilai seberapa baik model mampu mendeteksi seluruh data positif yang sebenarnya ada. F1-score, sebagai rata-rata harmonis antara presisi dan recall, memberikan gambaran yang seimbang, khususnya saat terdapat ketidakseimbangan kelas pada data dan akan ditampilkan dalam bentuk confusion matrix, dimana penggunaan confusion matrix telah terbukti efektif dalam mengukur kinerja model klasifikasi dan telah digunakan pada studi serupa [13]. Evaluasi ini tidak hanya digunakan untuk mengukur akurasi keseluruhan, tetapi juga untuk memahami *trade-off* antara performa dan kompleksitas model, terutama dalam konteks risiko kredit yang memiliki konsekuensi langsung pada keputusan bisnis. Evaluasi menyeluruh terhadap masing-masing algoritma sangat penting, karena performa model pembelajaran mesin dapat sangat bervaruasi tergantung pada struktur data dan metrik evaluasi yang digunakan [14].

### 2.6. Analisis Hasil dan Pembahasan

Tahap terakhir adalah analisis hasil dan pembahasan. Hasil evaluasi kinerja kedua model dibandingkan untuk menemukan model yang paling unggul dalam memprediksi kelayakan kredit pada sistem pinjaman online. Perbandingan dilakukan dengan melihat nilai akurasi, presisi, recall, dan F1-score, sehingga diperoleh pemahaman yang menyeluruh mengenai kelebihan masing-masing algoritma.

Dalam penelitian ini, perangkat lunak yang digunakan adalah Python sebagai bahasa pemrograman utama, dengan pustaka (*library*) pendukung seperti Scikit-learn untuk implementasi model, Pandas dan NumPy untuk manipulasi data, serta Matplotlib untuk visualisasi hasil. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem penilaian risiko kredit berbasis data yang lebih akurat dan efisien.

## 3. HASIL PENELITIAN DAN PEMBAHASAN

### 3.1. Hasil Pengolahan Data

Dataset yang digunakan dalam penelitian ini adalah *Loan Approval Prediction Dataset* yang diperoleh dari platform Kaggle. Dataset ini mencakup berbagai atribut penting yang merepresentasikan profil peminjam, seperti jumlah tanggungan, status pendidikan, status pekerjaan, pendapatan tahunan, jumlah pinjaman, jangka waktu pinjaman, skor CIBIL, nilai aset residential, nilai aset bank, dan status kelayakan pinjaman. Data ini sangat relevan untuk mengevaluasi faktor-faktor yang mempengaruhi kelayakan kredit pada layanan pinjaman online.

Sebelum dilakukan tahap pemodelan, dataset ini melalui proses pembersihan dan normalisasi. Proses pembersihan data mencakup penghapusan data duplikat, pengisian nilai kosong yang relevan, serta identifikasi dan penghapusan data yang dapat mempengaruhi kualitas model. Berdasarkan hasil analisis, tidak ditemukan data duplikat maupun nilai kosong pada dataset, tetapi terdapat 28 nilai yang tidak wajar pada atribut residential_assets_value, yaitu sebesar -100.000, di mana aset tidak seharusnya bernilai negatif, maka data tersebut akan dihapus dari dataset untuk mencegah potensi gangguan terhadap proses pelatihan model. Selanjutnya, data kategorikal seperti status *education* dan *self_employed* dikodekan menggunakan metode label *encoding* agar dapat diinterpretasikan sebagai data numerik oleh algoritma pembelajaran mesin.

Setelah tahap pra-pemrosesan, dataset dibagi menjadi 80% data latih dan 20% data uji menggunakan metode *train-test split*. Langkah ini penting untuk memastikan bahwa evaluasi kinerja model dilakukan secara objektif dan akurat, karena data uji tidak pernah digunakan pada tahap pelatihan model. Data yang telah disiapkan untuk proses dan pengujian dapat dilihat pada Tabel 1 berikut ini:

**Tabel 1.** Dataset

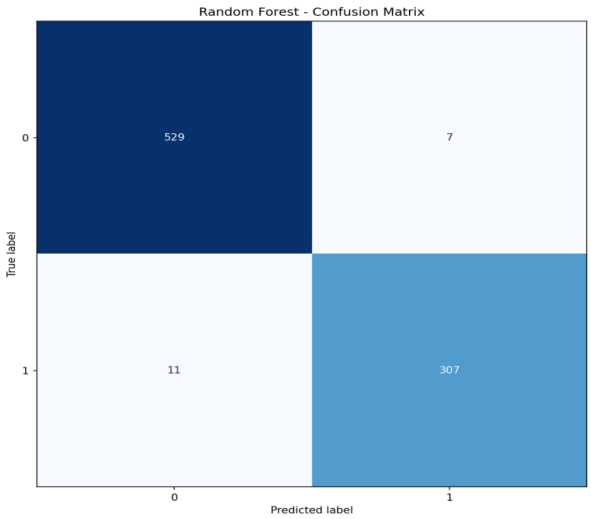
| No | Dependents | Education | ... | Employment Status | Loan Status |
|---|---|---|---|---|---|
| 1 | 2 | Graduate | ... | No | 1 |
| 2 | 0 | Not Graduate | ... | Yes | 0 |
| 3 | 3 | Graduate | ... | No | 0 |
| 4 | 3 | Graduate | ... | No | 0 |
| 5 | 5 | Not Graduate | ... | Yes | 0 |
| ... | ... | ... | ... | ... | ... |
| 4269 | 1 | Not Graduate | ... | No | 1 |
| 4270 | 1 | Graduate | ... | No | 1 |

### 3.2. Hasil Kinerja Model

Penelitian ini menggunakan dua algoritma klasifikasi, yaitu Random Forest dan KNN, untuk memprediksi kelayakan kredit pada data pinjaman online. Model KNN menggunakan nilai *k* = 5, yang merupakan hasil uji coba dengan pendekatan *grid search* sederhana, di mana nilai *k* = 5 memberikan akurasi terbaik pada data latih (tidak *overfitting*, stabil, dan akurat). Nilai k yang terlalu kecil berisiko menyebabkan *overfitting*, sedangkan nilai yang terlalu besar dapat mengurangi sensitivitas terhadap pola lokal.

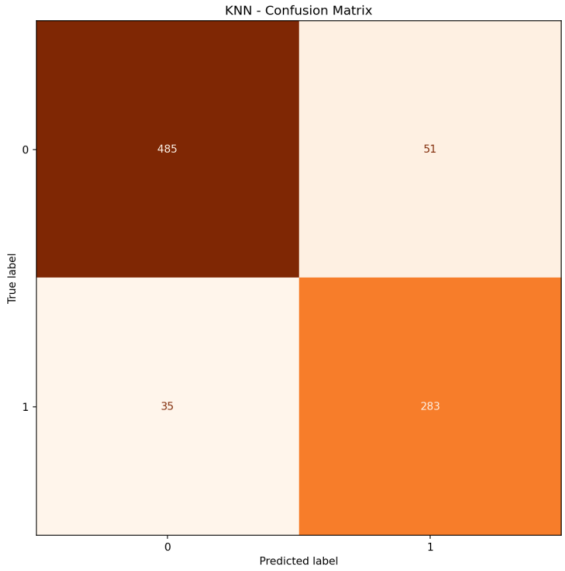
Model Random Forest dibangun dengan parameter *n_estimators* = 100, yang berarti model Random Forest akan membentuk 100 pohon keputusan yang independen. Setiap pohon dilatih pada subset data yang berbeda, dan prediksi akhir ditentukan melalui proses voting

mayoritas dari keseluruhan pohon. Jumlah 100 dipilih karena sudah cukup untuk memastikan stabilitas hasil tanpa menyebabkan komputasi yang terlalu berat.



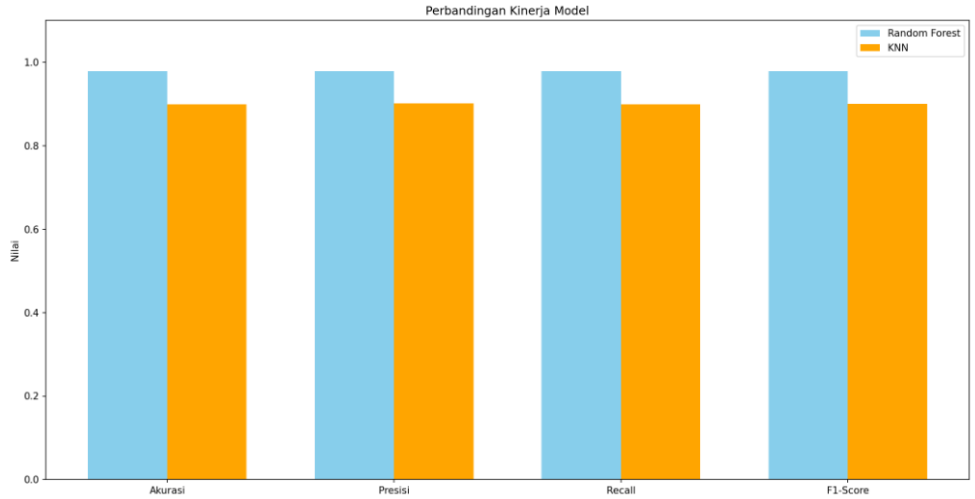

**Gambar 2.** *Confusion matrix* Random Forest

Berdasarkan hasil *confusion matrix* dari model Random Forest yang ditampilkan pada Gambar 2, dapat disimpulkan bahwa algoritma ini memiliki performa yang sangat baik dalam memprediksi kelayakan kredit. Dari total 854 data uji, model berhasil mengklasifikasikan 529 data *Rejected* dengan benar sebagai True Negative, dan 307 data *Approved* dengan benar sebagai True Positive. Hanya terdapat 7 data *Rejected* yang salah diklasifikasikan sebagai *Approved* (False Positive), serta 11 data *Approved* yang salah diklasifikasikan sebagai *Rejected* (False Negative). Hasil ini menunjukkan bahwa model Random Forest memiliki sensitivitas (recall terhadap kelas *Approved*) dan spesifisitas (kemampuan mendeteksi kelas *Rejected*) yang sangat baik, serta tingkat kesalahan klasifikasi yang sangat rendah. Hasil ini memperkuat posisi algoritma ini sebagai salah satu model klasifikasi terbaik dalam tugas prediksi risiko kredit, sebagaimana didukung oleh penelitian sebelumnya [15].




**Gambar 3.** *Confusion matrix* KNN

Berdasarkan hasil *confusion matrix* untuk model KNN yang ditampilkan pada Gambar 3, algoritma ini menunjukkan performa yang cukup baik dalam memprediksi kelayakan kredit. Dari total 854 data uji, model KNN berhasil mengklasifikasikan 485 data *Rejected* dengan benar sebagai True Negative, dan 283 data *Approved* dengan benar sebagai True Positive. Namun demikian, terdapat 51 data *Rejected* yang salah diprediksi sebagai *Approved* (False Positive), dan 35 data *Approved* yang salah diklasifikasikan sebagai *Rejected* (False Negative). Meskipun performanya tidak sebaik Random Forest, model KNN tetap menunjukkan tingkat akurasi yang cukup tinggi dan dapat dijadikan alternatif untuk aplikasi yang memerlukan algoritma yang lebih sederhana dan cepat dari segi komputasi.




**Gambar 4.** Diagram *Random Forest* dan *K-Nearest Neighbor*

Berdasarkan diagram batang yang ditampilkan pada Gambar 4, terlihat dengan jelas perbandingan kinerja antara model Random Forest dan KNN pada empat metrik evaluasi utama, yaitu akurasi, presisi, recall, dan F1-score. Warna biru muda digunakan untuk mewakili kinerja Random Forest, sedangkan warna jingga mewakili KNN.

Dari grafik tersebut, dapat diamati bahwa Random Forest menunjukkan performa yang lebih tinggi dibandingkan KNN pada keempat metrik. Meskipun selisihnya tidak terlalu besar, Random Forest secara konsisten unggul tipis dalam hal akurasi, presisi, recall, maupun F1-score. Hal ini mengindikasikan bahwa Random Forest lebih andal dalam mengenali pola pada data kelayakan kredit dibandingkan dengan KNN.

**Tabel 2.** Evaluasi Kinerja Model

| Model | Akurasi | Presisi | *Recall* | *F1-score* |
|---|---|---|---|---|
| Random Forest | 97% | 97% | 97% | 97% |
| KNN | 89% | 90% | 89% | 89% |

Berdasarkan hasil evaluasi yang ditampilkan pada Tabel 2, terlihat perbandingan kinerja antara model Random Forest dan KNN. Model Random Forest menunjukkan performa yang sangat baik dan konsisten pada seluruh metrik evaluasi, dengan nilai sebesar 97% untuk akurasi, presisi, recall, dan F1-score. Hasil ini mencerminkan stabilitas serta kemampuan model dalam mempelajari dan memprediksi data secara efektif. Sementara itu, model KNN memperoleh nilai akurasi sebesar 89%, dengan presisi sedikit lebih tinggi, yaitu 90%. Meskipun nilai recall dan F1-score pada KNN lebih rendah dibandingkan dengan Random Forest, model ini tetap

menunjukkan kinerja yang cukup baik secara keseluruhan dan dapat dipertimbangkan sebagai alternatif yang efisien pada kondisi tertentu.

## 4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa algoritma Random Forest menunjukkan performa yang lebih unggul dibandingkan K-Nearest Neighbor (KNN) dalam tugas prediksi kelayakan kredit pada sistem pinjaman online. Hal ini ditunjukkan oleh nilai akurasi, presisi, recall, dan F1-score yang lebih tinggi dan konsisten di semua metrik evaluasi. Random Forest memiliki keunggulan dalam mengatasi data yang kompleks, sedangkan KNN tetap menunjukkan kinerja yang baik meskipun lebih sederhana. Secara praktis, penelitian ini dapat digunakan sebagai acuan dalam pengembangan sistem pendukung keputusan untuk penyedia layanan pinjaman online, khususnya dalam menyaring risiko kredit calon peminjam secara otomatis dan cepat. Dari sisi akademik, penelitian ini memberikan kontribusi terhadap pengembangan studi komparatif antara algoritma pembelajaran mesin dalam konteks prediksi risiko keuangan.

Hasil penelitian ini juga menunjukkan bahwa KNN masih layak dipertimbangkan sebagai model alternatif dalam klasifikasi kelayakan kredit, terutama untuk aplikasi dengan kebutuhan komputasi yang lebih ringan. Oleh karena itu saran yang dapat diberikan untuk penelitian selanjutnya adalah mencoba memadukan algoritma Random Forest dan KNN dalam model *ensemble* atau *hybrid*, serta mempertimbangkan penggunaan data yang lebih besar dan beragam untuk memperoleh hasil yang lebih komprehensif dan generalisasi yang lebih baik.

## PERNYATAAN PENGGUNAAN ARTIFICIAL INTELLIGENCE

Penulis menggunakan DeepSeek dan Gemini sebagai alat bantu untuk mendukung penyelesaian aspek teknis penelitian dan penyempurnaan penulisan artikel. Saran terkait pemrograman, pengorganisasian data, dan penyajian informasi diperiksa kembali sebelum digunakan. Penulis memastikan bahwa seluruh hasil yang dilaporkan sesuai dengan data penelitian dan bertanggung jawab penuh atas isi publikasi.

## KONTRIBUSI PENULIS

**Santi Prayudani:** Konsep, Metode, Analisis Data, Penulisan Artikel, Supervisi, Review.

**Yous Sibarani:** Pemrograman, Analisis Data, Penulisan Artikel.

**Azrizal Salam:** Metode, Analisis Data, Penulisan Artikel

**Arif Ridho Lubis:** Analisis Data, Penulisan Artikel, Pemrograman, Editing].

## PENDANAAN

Penelitian ini tidak menerima hibah khusus dari lembaga pendanaan mana pun, baik sektor publik, komersial, maupun nirlaba.

## KONFLIK KEPENTINGAN

Penulis menyatakan tidak terdapat konflik kepentingan terkait publikasi artikel ini.

## DAFTAR PUSTAKA

[1] A. Abdullah, “Analisis Pengetahuan Pinjaman Online Pada Masyarakat Surakarta,” JESI (Jurnal Ekonomi Syariah Indonesia), vol. 11, no. 2, pp. 108–114, 2021, doi: 10.21927/jesi.2021.11(2).108-114.

[2] U. Makhazhanova, S. Kerimkhulle, A. Mukhanova, A. Bayegizova, Z. Aitkozha, A. Mukhiyadin, B. Tassuov, A. Saliyeva, R. Taberkhan, and G. Azieva, "The evaluation of creditworthiness of trade and enterprises of service using the method based on fuzzy logic," Applied Sciences, vol. 12, no. 22, p. 11515, 2022, doi: 10.3390/app122211515.
[3] M. A. Faheem, "AI-Driven Risk Assessment Models: Revolutionizing Credit Scoring and Default Prediction," Iconic Research and Engineering Journals, vol. 5, no. 3, pp. 177– 186, 2021. [Online]. Available: https://www.irejournals.com/formatedpaper/1702907.pdf
[4] A. Ampountolas, T. Nyarko Nde, P. Date, and C. Constantinescu, "A machine learning approach for micro-credit scoring," Risks, vol. 9, no. 3, p. 50, 2021, doi: 10.3390/risks9030050.
[5] A. Kumar, S. Sharma, and M. Mahdavi, "Machine learning (ML) technologies for digital credit scoring in rural finance: A literature review," Risks, vol. 9, no. 11, p. 192, 2021, doi: 10.3390/risks9110192.
[6] V. Chang, Q. A. Xu, S. H. Akinloye, V. Benson, and K. Hall, "Prediction of bank credit worthiness through credit risk analysis: An explainable machine learning study," Annals of Operations Research, advance online publication, 2024, doi: 10.1007/s10479-024-06134-x.
[7] E. V. Orlova, "Methodology and Models for Individuals' Creditworthiness Management Using Digital Footprint Data and Machine Learning Methods," Mathematics, vol. 9, no. 15, p. 1820, 2021, doi: 10.3390/math9151820.
[8] D. Fusvita, A. Asnawati, and F. H. Utami, "Penerapan algoritma KNN (K-nearest neighbour) dalam klasifikasi data pinjaman anggota koperasi," J. Ilm. Binary STMIK Bina Nusantara Jaya, vol. 3, no. 1, pp. 1–5, Apr. 2021, doi: 10.52303/jb.v3i1.32.
[9] S. Shi, R. Tse, W. Luo, S. D'Addona, and G. Pau, "Machine learning-driven credit risk: a systemic review," Neural Computing and Applications, vol. 34, pp. 14327–14339, 2022, doi: 10.1007/s00521-022-07472-2.
[10] D. Mhlanga, "Financial Inclusion in Emerging Economies: The Application of Machine Learning and Artificial Intelligence in Credit Risk Assessment," International Journal of Financial Studies, vol. 9, no. 3, 2021, doi: 10.3390/ijfs9030039.
[11] B. Prasojo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," TEKNOSI, vol. 7, no. 2, pp. 79–89, 2021, doi: 10.25077/TEKNOSI.v7i2.2021.79-89.
[12] A. Y. Vandika and R. Pannyiwi, "Application of KNN Algorithm for Credit Risk Analysis in Savings and Loan Cooperatives," JICT, vol. 15, no. 2, pp. 55–61, 2024, doi: 10.35335/jict.v15i2.174. [Online]. Available: https://www.ejournal.marqchainstitute.or.id/index.php/JICT/article/view/174
[13] S. G. Wardhani and A. Kurniawati, "Implementation of K-Nearest Neighbor Algorithm for Creditworthiness Analysis Using Methods Cross-Industry Standard Process for Data Mining (CRISP-DM)," Int. Res. J. Adv. Eng. Sci., vol. 10, no. 1, pp. 152–157, 2025. [Online]. Available: https://irjaes.com/wp-content/uploads/2025/02/IRJAES-V10N1P420Y25.pdf
[14] H. Dong, R. Liu, and A. W. Tham, "Accuracy Comparison between Five Machine Learning Algorithms for Financial Risk Evaluation," Journal of Risk and Financial Management, vol. 17, no. 2, 2024, doi: 10.3390/jrfm17020050.
[15] M. R. Machado and S. Karray, "Assessing credit risk of commercial customers using hybrid machine learning algorithms," Expert Systems with Applications, vol. 200, p. 116889, 2022, doi: 10.1016/j.eswa.2022.116889.