

# Penerapan Algoritma Random Forest dalam Klasifikasi Persetujuan Pinjaman Menggunakan Dataset Loan Approval

Afiq Muhammad Nur Marzuki<sup>1\*</sup>, Devan Vidhel Azadan<sup>2</sup>, Atik Setyaningih<sup>3</sup>, Satria Abdul Rofiq<sup>4</sup>

<sup>1</sup>Teknik Informasi/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

<sup>1\*</sup>230103045@mhs.udb.ac.id

<sup>2</sup>Teknik Informasi/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

<sup>2</sup>230103055@mhs.udb.ac.id

<sup>3</sup>Teknik Informasi/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

<sup>3</sup>230103052@mhs.udb.ac.id

<sup>4</sup>Teknik Informasi/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta)

<sup>4</sup>230103077@mhs.udb.ac.id

**Abstrak**— Operasional lembaga keuangan sangat bergantung pada ketepatan proses persetujuan pinjaman, yang memegang peran sentral dalam penilaian kelayakan kredit serta mitigasi risiko. Studi ini berfokus pada implementasi algoritma Random Forest guna mengkategorikan status persetujuan kredit melalui Loan Approval Dataset. Kami menerapkan kerangka kerja Knowledge Discovery in Database (KDD) yang mencakup tahap seleksi, prapemrosesan, transformasi, penambahan data, serta evaluasi. Data yang diproses melibatkan lima atribut prediktor yaitu Age, Gender, Income, LoanAmount, dan CreditScore, dengan Loan\_Status sebagai variabel target yang mencakup klasifikasi Approved serta Rejected. Model dikembangkan melalui algoritma Random Forest dengan pembagian komposisi 80% untuk pelatihan dan 20% untuk pengujian. Evaluasi kinerja menghasilkan Accuracy 95,00%, Precision 95,89%, Recall 97,22%, serta F1-Score 96,55%, yang mengindikasikan tingkat performa klasifikasi yang optimal. Analisis Feature Importance mengungkapkan bahwa CreditScore adalah variabel paling dominan yang memengaruhi keputusan kredit, disusul oleh Income dan LoanAmount. Lebih lanjut, model ini telah diintegrasikan ke dalam sistem berbasis Streamlit, yang berfungsi sebagai perangkat pendukung keputusan untuk mempercepat analisis persetujuan pinjaman secara objektif serta efisien.

**Kata kunci**— Data Mining, Random Forest, Klasifikasi, Loan Approval, Machine Learning.

**Abstract**— The operations of financial institutions are highly dependent on the accuracy of the loan approval process, which plays a central role in assessing creditworthiness and mitigating risks. This study focuses on the implementation of the Random Forest algorithm to categorize loan approval statuses using the Loan Approval Dataset. We applied the Knowledge Discovery in Database (KDD) framework, which includes selection, preprocessing, transformation, data mining, and evaluation stages. The processed data involved five predictor attributes: Age, Gender, Income, LoanAmount, and CreditScore, with Loan\_Status as the target variable, encompassing both Approved and Rejected classifications. The model was developed using the Random Forest algorithm with a composition split of 80% for training and 20% for testing. The performance evaluation yielded an Accuracy of 95.00%, Precision of 95.89%, Recall of 97.22%, and F1-Score of 96.55%, indicating optimal classification performance. Feature Importance analysis revealed that CreditScore was the most dominant variable influencing credit decisions, followed by Income and LoanAmount. Furthermore, this model has been integrated into a Streamlit-based system, which serves as a decision support tool to accelerate loan approval analysis objectively and efficiently.

**Keywords**— Data Mining, Random Forest, Classification, Loan Approval, Machine Learning.

## I. PENDAHULUAN

Perkembangan Data Mining dan Machine Learning memberikan dampak nyata terhadap optimalisasi proses pengambilan keputusan di berbagai ranah, dengan fokus utama pada sektor keuangan. Lembaga keuangan seperti bank, koperasi, dan perusahaan Financial Technology (FinTech) memanfaatkan teknik analisis data

untuk membantu proses evaluasi kelayakan pinjaman secara lebih cepat, objektif, dan akurat. Proses persetujuan pinjaman (loan approval) merupakan tahapan penting karena berkaitan langsung dengan risiko kredit yang dapat memengaruhi stabilitas keuangan lembaga. Kesalahan dalam menentukan kelayakan pemohon berpotensi meningkatkan jumlah

kredit bermasalah (non-performing loan) sehingga diperlukan sistem klasifikasi yang mampu menghasilkan keputusan secara konsisten berdasarkan data historis [1], [2],[3].

Pada praktiknya, proses analisis kelayakan pinjaman masih banyak dilakukan secara manual melalui pemeriksaan dokumen dan penilaian oleh analis kredit. Pendekatan tersebut memerlukan waktu yang relatif lama, rentan terhadap subjektivitas, serta kurang efisien ketika jumlah pengajuan pinjaman terus meningkat. Selain itu, karakteristik setiap pemohon yang berbeda, seperti usia, jenis kelamin, pendapatan, jumlah pinjaman, dan skor kredit, menyebabkan proses evaluasi menjadi semakin kompleks. Oleh karena itu, penerapan metode klasifikasi berbasis Data Mining diperlukan untuk membantu lembaga keuangan dalam mengidentifikasi pola data dan menghasilkan keputusan persetujuan pinjaman secara lebih efektif dan akurat [2], [4] , [3], [5], [6].

Salah satu algoritma klasifikasi yang banyak digunakan adalah Random Forest, yaitu metode ensemble learning yang membangun sejumlah decision tree menggunakan teknik bootstrap sampling dan bagging, kemudian menentukan hasil klasifikasi melalui mekanisme majority voting. Dibandingkan algoritma berbasis pohon keputusan tunggal, Random Forest memiliki kemampuan yang lebih baik dalam mengurangi overfitting, menangani data yang kompleks, serta menghasilkan performa klasifikasi yang lebih stabil [7]. Selain itu, algoritma ini mampu mengukur tingkat kepentingan setiap atribut (feature importance) sehingga dapat diketahui faktor-faktor yang paling berpengaruh terhadap hasil klasifikasi [1], [8], [3], [9].

Beberapa penelitian terdahulu menunjukkan bahwa Random Forest memiliki performa yang baik dalam klasifikasi kelayakan kredit maupun persetujuan pinjaman. Andi dan Thamrin [1] membuktikan bahwa Random Forest memberikan performa yang lebih baik dibandingkan Support Vector Machine dalam analisis kelayakan kredit nasabah. Rahmawati dan Handayani [2] berhasil menerapkan Random Forest untuk klasifikasi kelayakan

kredit pada PT BPR BKK Wonogiri sehingga mampu mendukung pengelolaan risiko keuangan. [8] meningkatkan performa prediksi persetujuan pinjaman melalui penerapan hyperparameter tuning, sedangkan [10] memperlihatkan bahwa Random Forest memberikan tingkat akurasi yang lebih unggul daripada Naïve Bayes dalam penentuan kelayakan kredit [11]. Temuan ini memperkuat argumen bahwa algoritma ini sangat kompeten dalam menangani berbagai tantangan klasifikasi pada sektor finansial.

Kendati demikian, literatur terdahulu cenderung memusatkan perhatian pada komparasi performa antar algoritma atau upaya optimasi akurasi model, sedangkan pembahasan mengenai kontribusi masing-masing atribut terhadap keputusan persetujuan pinjaman masih terbatas. Di sisi lain, pemanfaatan model klasifikasi ke dalam antarmuka web yang intuitif bagi pengguna masih tergolong jarang dikembangkan [8],[12][3]. Berdasarkan kondisi tersebut, studi ini mengadopsi algoritma Random Forest guna mengkategorikan status kelayakan kredit pinjaman menggunakan Loan Approval Dataset yang terdiri atas atribut Age, Gender, Income, LoanAmount, dan CreditScore dengan target klasifikasi Loan\_Status. Penelitian dilakukan menggunakan Proses pengolahan data kami berpedoman pada kerangka kerja Knowledge Discovery in Database (KDD), yang mencakup tahapan seleksi, preprocessing, transformasi, penambangan data, serta evaluasi. Proses evaluasi model divalidasi melalui Confusion Matrix dengan memantau metrik Accuracy, Precision, Recall, serta F1-Score. Selain itu, kami menyertakan analisis Feature Importance untuk memetakan atribut mana yang memiliki kontribusi paling signifikan terhadap keputusan persetujuan pinjaman. Hasil penelitian diharapkan dapat menghasilkan model klasifikasi yang akurat sekaligus memberikan informasi yang bermanfaat bagi lembaga keuangan selain itu, implementasi model klasifikasi ke dalam antarmuka web yang intuitif bertujuan agar langkah pengambilan keputusan menjadi lebih presisi serta mampu mempercepat alur kerja organisasi [13], [14]

## II. METODOLOGI PENELITIAN

Penelitian ini menerapkan metode Data Mining untuk memproses Loan Approval Dataset yang mencakup 1.000 entri data, guna mentransformasinya menjadi wawasan strategis dalam mendukung pengambilan keputusan persetujuan pinjaman. Pengembangan model klasifikasi ini diimplementasikan melalui metodologi Knowledge Discovery in Database (KDD) yang memberikan alur kerja terstruktur, mencakup seluruh tahapan mulai dari seleksi data hingga evaluasi performa model secara komprehensif.

Melalui implementasi algoritma Random Forest, riset ini berfokus pada konstruksi model klasifikasi yang mampu membedakan calon peminjam yang layak (Approved) dan tidak layak (Rejected) menerima pinjaman berdasarkan karakteristik yang dimiliki. Model yang dihasilkan selanjutnya diimplementasikan ke dalam aplikasi berbasis Streamlit sehingga proses prediksi status persetujuan pinjaman dapat dilakukan secara interaktif dan menunjang pengambilan keputusan yang lebih akurat, terukur, serta berbasis data.

### A. Metode Penelitian

Penelitian ini menerapkan metodologi kuantitatif untuk membangun model klasifikasi persetujuan pinjaman berbasis algoritma Random Forest. Pendekatan kuantitatif diterapkan karena riset ini menitikberatkan pada analisis numerik guna memodelkan klasifikasi status pinjaman berdasarkan profil pemohon. Riset ini mengadopsi alur kerja Knowledge Discovery in Database (KDD) guna membangun model klasifikasi. Proses ini mencakup lima tahap utama: seleksi, pra-pemrosesan, transformasi, ekstraksi data, serta evaluasi [12],[8].

### B. Sumber Data

Data dianalisis dalam riset ini merujuk pada Loan Approval Dataset yang diakses melalui platform Kaggle [12]. Kumpulan data tersebut memuat 1.000 entri dengan tujuh atribut spesifik, yaitu Applicant\_ID, Age, Gender, Income,

LoanAmount, CreditScore, dan Loan\_Status. Variabel Loan\_Status digunakan sebagai variabel target yang terdiri atas dua kelas, yaitu Approved (disetujui) sebanyak 503 data dan Rejected (tidak disetujui) sebanyak 497 data. Sementara itu, atribut Applicant\_ID hanya berfungsi sebagai identitas data sehingga tidak digunakan dalam proses pembentukan model klasifikasi.

Berdasarkan dataset tersebut, atribut Age, Gender, Income, LoanAmount, dan CreditScore digunakan sebagai variabel prediktor (independent variable), sedangkan Loan\_Status digunakan sebagai variabel target (dependent variable) dalam proses klasifikasi menggunakan algoritma Random Forest.

Rincian atribut yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Potensi Konversi Beberapa Radionuklida

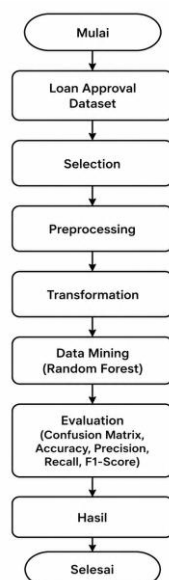
| No | Nama Variabel | Tipe Data   | Jenis Variabel | Keterangan                                  |
|----|---------------|-------------|----------------|---------------------------------------------|
| 1  | Applicant_ID  | Integer     | Identitas      | Nomor unik identifikasi nasabah             |
| 2  | Age           | Integer     | Independen     | Usia pemohon dalam satuan tahun             |
| 3  | Gender        | Categorical | Independen     | Jenis kelamin nasabah (Laki-laki/Perempuan) |
| 4  | Income        | Numeric     | Independen     | Total pendapatan tahunan pemohon            |
| 5  | LoanAmount    | Numeric     | Independen     | Total nominal pinjaman yang diajukan        |
| 6  | CreditScore   | Numeric     | Independen     | Nilai riwayat kredit nasabah                |
| 7  | Loan_Status   | Categorical | Dependen       | Status akhir keputusan (Approved/Rejected)  |

Berdasarkan Tabel 1, Loan\_Status merupakan variabel target yang diprediksi, sedangkan Age, Gender, Income, LoanAmount, dan CreditScore adalah variabel independen (fitur) sebagai dasar pelatihan model. Applicant\_ID dikecualikan dari proses pelatihan karena hanya berfungsi sebagai identitas unik nasabah.

### C. Tahapan Data Mining

Riset ini mengadopsi kerangka kerja Knowledge Discovery in Database (KDD) guna menjamin alur pengolahan data yang sistematis dan terstruktur dalam membangun model klasifikasi persetujuan pinjaman. Metode KDD dipilih karena mampu mengintegrasikan seluruh tahapan pengolahan data, mulai dari pemilihan data hingga evaluasi model, sehingga menghasilkan klasifikasi yang valid sebagai basis acuan dalam penilaian kelayakan pemohon [12], [8]. Seluruh tahapan metodologi penelitian ini dirangkum dalam Gambar 1.

Secara umum, tahapan penelitian dimulai dari proses Selection untuk memilih atribut yang relevan, dilanjutkan dengan Preprocessing guna meningkatkan kualitas data melalui pemeriksaan dan pembersihan data. Pada fase Transformation, data diproses kembali agar formatnya selaras dengan kebutuhan algoritma klasifikasi. Pada tahap Data Mining, algoritma Random Forest diterapkan untuk menyusun model prediksi terkait persetujuan pinjaman. Tahap terakhir adalah Evaluation, yaitu Performa model diuji melalui Confusion Matrix untuk mendapatkan nilai Accuracy, Precision, Recall, serta F1-Score. Melalui metrik ini, model klasifikasi dipastikan mampu memberikan prediksi persetujuan pinjaman yang andal dan terukur.



Gambar 1. Alur Metodologi Penelitian

#### 1) Selection

Tahap Selection dilakukan dengan memilih atribut data yang paling relevan untuk menunjang performa model. Pada tahap ini dilakukan identifikasi terhadap variabel yang memiliki relevansi dengan tujuan penelitian, yaitu mengklasifikasikan status persetujuan pinjaman. Merujuk pada Tabel 1, atribut Applicant\_ID dieliminasi karena hanya berperan sebagai pengenal data unik dan tidak memiliki nilai informatif dalam proses klasifikasi. Oleh karena itu, atribut yang digunakan sebagai variabel prediktor meliputi Age, Gender, Income, LoanAmount, dan CreditScore, sedangkan Loan\_Status digunakan sebagai variabel target.

Pemilihan atribut tersebut didasarkan pada relevansinya terhadap karakteristik pemohon beserta variabel-variabel penentu yang menjadi acuan utama dalam evaluasi kredit. Atribut Age dan Gender merepresentasikan karakteristik demografis pemohon, sedangkan Income, LoanAmount, dan CreditScore menggambarkan kondisi finansial serta kemampuan pemohon dalam memenuhi kewajiban pembayaran pinjaman. Dengan menggunakan atribut-atribut tersebut, model diharapkan mampu mempelajari pola yang membedakan karakteristik pemohon yang memperoleh status Approved dan Rejected, sehingga proses klasifikasi dapat dilakukan secara lebih akurat.

#### 2) Prerprocessing

Tahap Preprocessing bertujuan memvalidasi serta memperbaiki kualitas data agar model klasifikasi yang dihasilkan nantinya lebih optimal. Proses tersebut mencakup serangkaian langkah krusial pemeriksaan terhadap missing value, data duplikat, konsistensi tipe data, serta validitas nilai pada setiap atribut. Proses preprocessing merupakan tahapan yang penting karena kualitas data akan memengaruhi kinerja model klasifikasi yang dihasilkan.

Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki missing value maupun data duplikat sehingga tidak diperlukan proses imputasi atau penghapusan data. Selanjutnya dilakukan pemeriksaan tipe data dan konsistensi nilai pada setiap atribut untuk memastikan

konsistensi format data agar memenuhi syarat teknis sebelum memasuki fase transformasi. Dengan demikian, dataset dinyatakan siap digunakan pada proses pembentukan model Random Forest.

### 3) Transformation

Tahap Transformation mengonversi data ke format yang kompatibel untuk diolah lebih lanjut oleh algoritma Random Forest. Atribut kategorikal dikonversi menjadi data numerik melalui Label Encoding agar memenuhi format input yang disyaratkan oleh model klasifikasi.

Proses transformasi dilakukan pada atribut Gender dan Loan\_Status. Atribut Gender dikonversi menjadi nilai numerik untuk merepresentasikan jenis kelamin pemohon, sedangkan Loan\_Status diubah menjadi label numerik sebagai variabel target yang nantinya diproses oleh model. Setelah tahap encoding selesai, dataset dipisahkan menjadi variabel prediktor (X) yang terdiri atas Age, Gender, Income, LoanAmount, dan CreditScore, serta variabel target (Y) yaitu Loan\_Status.

Fase transformasi ini krusial untuk memastikan seluruh data memiliki format konsisten yang menjadi prasyarat bagi pemrosesan algoritma konsisten sehingga proses pembentukan model Random Forest dapat dilakukan secara optimal tanpa mengalami kendala akibat perbedaan tipe data.

Tabel 2. Hasil Transformasi Data

| Atribut     | Sebelum Transformasi  | Setelah Transformasi          |
|-------------|-----------------------|-------------------------------|
| Gender      | Male, Female          | Male = 1,<br>Female = 2       |
| Loan Status | Approved,<br>Rejected | Approved = 1,<br>Rejected = 0 |

### 4) Data Mining

Tahap ini menjadi inti penelitian, di mana model klasifikasi dibentuk menggunakan algoritma Random Forest Classifier. Sebelum pelatihan model, dataset dipartisi menjadi 80% data pelatihan dan 20% data pengujian dengan metode Hold-Out Validation. Skema pembagian ini dilakukan agar model dapat mengenali pola pada data latih sekaligus menguji kinerjanya

pada data yang belum pernah diproses sebelumnya, sehingga evaluasi kemampuan prediksi menjadi lebih objektif.

Pada tahap pelatihan, algoritma Random Forest membangun sejumlah Decision Tree berdasarkan data hasil bootstrap sampling, yaitu pengambilan sampel dilakukan secara acak melalui metode sampling with replacement. Di samping itu, tiap Decision Tree dikembangkan secara independen menggunakan subset data yang berbeda dibentuk menggunakan kombinasi atribut yang dipilih secara acak (random feature selection) sehingga setiap pohon memiliki karakteristik yang berbeda. Selanjutnya, hasil prediksi dari seluruh Decision Tree digabungkan menggunakan mekanisme majority voting untuk menentukan kelas akhir, yaitu Approved atau Rejected [1], [2], [8].

Penerapan algoritma Random Forest dipilih karena mampu menangani data dengan karakteristik yang beragam, memiliki ketahanan terhadap overfitting, serta memberikan performa klasifikasi yang baik pada permasalahan prediksi kelayakan kredit dan persetujuan pinjaman. Selain menghasilkan prediksi, algoritma ini juga mampu memberikan nilai Feature Importance yang digunakan untuk mengidentifikasi atribut yang paling berpengaruh terhadap keputusan persetujuan pinjaman. Informasi tersebut menjadi salah satu dasar dalam melakukan analisis faktor-faktor yang memengaruhi hasil klasifikasi pada penelitian ini.

### 5) Evaluation

Tahap Evaluation dilakukan untuk mengukur dan menganalisis performa model klasifikasi Random Forest dalam memprediksi status persetujuan pinjaman. Evaluasi bertujuan untuk mengetahui seberapa akurat model dalam memetakan data ke dalam kategori Approved maupun Rejected menggunakan testing data. Tahap evaluasi ini mengandalkan Confusion Matrix sebagai acuan utama dalam mengukur performa model [8], [12].

Confusion Matrix menyajikan empat komponen dasar: True Positive (TP), True Negative (TN), False Positive (FP), dan False

Negative (FN). Kumpulan metrik ini kemudian menjadi landasan untuk mengalkulasi Accuracy, Precision, Recall, serta F1-Score, kemudian diuji menggunakan data yang belum pernah diproses sebelumnya sebagai tolok ukur objektif untuk menilai kinerja prediktif model. Penggunaan beberapa metrik evaluasi memberikan gambaran yang lebih objektif mengenai kualitas model dibandingkan hanya menggunakan nilai akurasi saja [8], [12]. Hasil evaluasi tersebut selanjutnya digunakan untuk menganalisis tingkat ketepatan model dalam mengklasifikasikan status persetujuan pinjaman serta menjadi dasar dalam menilai kelayakan algoritma Random Forest untuk diterapkan pada permasalahan klasifikasi persetujuan pinjaman.

#### D. Algoritma Random Forest

Random Forest merupakan salah satu algoritma klasifikasi berbasis ensemble learning yang bekerja dengan membangun sejumlah Decision Tree kemudian menggabungkan hasil prediksi dari setiap pohon keputusan untuk menghasilkan keputusan akhir yang lebih akurat. Berbeda dengan Decision Tree tunggal, Random Forest memanfaatkan teknik bootstrap sampling dan random feature selection sehingga setiap pohon keputusan dibangun menggunakan sampel data dan kombinasi atribut yang berbeda. Pendekatan tersebut mampu meningkatkan kemampuan generalisasi model serta mengurangi risiko overfitting [1], [3], [14].

Pada penelitian ini, algoritma Random Forest Classifier digunakan untuk mengklasifikasikan status persetujuan pinjaman berdasarkan atribut Age, Gender, Income, LoanAmount, dan CreditScore. Selama proses pelatihan (training), algoritma membentuk sejumlah Decision Tree menggunakan data hasil bootstrap sampling. Setiap pohon kemudian menghasilkan prediksi terhadap status pinjaman, yaitu Approved atau Rejected. Selanjutnya, seluruh hasil prediksi digabungkan menggunakan mekanisme majority voting, sehingga kelas dengan perolehan suara terbanyak ditetapkan sebagai hasil prediksi akhir.

Secara matematis, proses pengambilan Keputusan pada algoritma Random Forest Dinyatakan sebagai berikut.

$$\hat{y} = \text{mode}\{T_i(x)\}_{i=1}^n$$

Keterangan:

- $\hat{Y}$  = hasil prediksi akhir.
- $T_1(x), T_2(x), \dots, T_n(x)$  = hasil prediksi dari masing-masing Decision Tree.
- $n$  = jumlah Decision Tree yang dibangun.
- mode** = kelas yang memiliki jumlah suara (majority voting) terbanyak

Selain menghasilkan prediksi klasifikasi, algoritma Random Forest juga menghitung Feature Importance, yaitu nilai yang menunjukkan tingkat kontribusi setiap atribut terhadap proses pengambilan keputusan model. Nilai tersebut dimanfaatkan dalam penelitian ini untuk menganalisis atribut yang paling berpengaruh terhadap persetujuan pinjaman, sehingga dapat memberikan informasi yang lebih komprehensif bagi proses analisis kelayakan kredit.

Pemilihan algoritma Random Forest didasarkan pada kemampuannya dalam menangani data numerik maupun kategorikal, menghasilkan akurasi klasifikasi yang tinggi, serta memiliki ketahanan terhadap noise dan overfitting. Selain itu, algoritma ini telah banyak digunakan pada penelitian mengenai klasifikasi kelayakan kredit dan persetujuan pinjaman dengan hasil yang baik, sehingga dinilai sesuai untuk diterapkan pada penelitian ini [1],[2],[8],[15],[10].

#### E. Evaluasi Model

Evaluasi model dilakukan untuk mengukur tingkat performa algoritma Random Forest dalam mengklasifikasikan status persetujuan pinjaman berdasarkan data pengujian (testing data). Kinerja model dievaluasi melalui Confusion Matrix yang menyajikan empat parameter dasar: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Keempat nilai tersebut kemudian diolah untuk memperoleh metrik Accuracy, Precision, Recall, serta F1-Score sebagai tolok ukur performa model [8], [12].

Confusion Matrix diterapkan sebagai instrumen pembandingan antara hasil prediksi model dengan data aktual guna mengukur tingkat akurasi klasifikasi. Dalam konteks ini, True Positive (TP) merujuk pada data Approved yang terprediksi dengan benar, sementara True Negative (TN) mencakup data Rejected yang terklasifikasi tepat. Sebaliknya, False Positive (FP) menandakan data Rejected yang keliru diprediksi sebagai Approved, dan False Negative (FN) merepresentasikan data Approved yang terklasifikasi sebagai Rejected.

Tingkat ketepatan model secara menyeluruh dihitung menggunakan metrik Accuracy, yang didefinisikan melalui persamaan berikut.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Metrik Precision mengkalkulasi tingkat ketepatan model dalam mengidentifikasi kelas Approved, dengan formula sebagai berikut.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Metrik Recall merepresentasikan cakupan model dalam mengidentifikasi seluruh data aktual kelas Approved, dengan formula berikut.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Performa model secara menyeluruh dihitung melalui F1-Score, yang menyelaraskan nilai Precision dan Recall ke dalam satu metrik tunggal.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

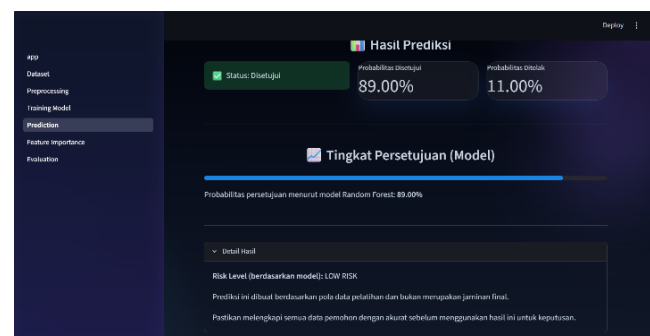
Penggunaan keempat metrik secara simultan ditujukan untuk memvalidasi performa Random Forest pada kasus klasifikasi pinjaman. Konsistensi skor tinggi pada setiap metrik menjadi indikator utama bahwa model memiliki tingkat reliabilitas yang memadai untuk diterapkan dalam proses persetujuan pinjaman.

### III. HASIL DAN PEMBAHASAN

#### A. Implementasi Sistem

Model klasifikasi yang telah dibangun menggunakan algoritma Random Forest kemudian diimplementasikan ke dalam aplikasi berbasis Streamlit sebagai sistem pendukung keputusan untuk memprediksi status persetujuan pinjaman. Sistem ini dirancang untuk memangkas durasi komputasi prediksi, di mana pengguna dapat memperoleh hasil klasifikasi secara instan tanpa perlu melakukan eksekusi pelatihan model secara berulang.

Sistem dirancang dengan antarmuka yang sederhana sehingga pengguna hanya perlu memasukkan data calon peminjam, meliputi Age, Gender, Income, Loan Amount, dan Credit Score. Tahap berikutnya melibatkan inferensi data menggunakan model Random Forest yang telah divalidasi guna menghasilkan keputusan berupa status Approved atau Rejected.



Gambar 2. Halaman Prediksi Persetujuan Pinjaman

Halaman prediksi merupakan halaman utama yang digunakan untuk melakukan klasifikasi status persetujuan pinjaman berdasarkan data calon peminjam. Pengguna memasukkan nilai Age, Gender, Income, LoanAmount, dan CreditScore, kemudian sistem akan memproses data tersebut menggunakan model Random Forest yang telah melalui tahap training.

Setelah tombol Prediksi ditekan, sistem menampilkan hasil klasifikasi berupa status Approved atau Rejected, disertai dengan nilai probabilitas persetujuan, probabilitas penolakan, tingkat persetujuan (approval probability), serta tingkat risiko (Risk Level) mengacu pada output inferensi model. Selanjutnya, sistem juga

memberikan informasi tambahan sebagai bahan pertimbangan untuk mendukung pengambilan keputusan.

Implementasi halaman prediksi mengindikasikan bahwa model Random Forest tidak hanya mampu memberikan performa klasifikasi yang optimal, melainkan juga dapat diimplementasikan ke dalam aplikasi berbasis Streamlit sebagai media pendukung pengambilan keputusan guna membantu proses analisis persetujuan pinjaman secara lebih cepat, objektif, dan efisien.

### B. Data Penelitian

Dataset yang dimanfaatkan pada penelitian ini berupa adalah Loan Approval Dataset yang terdiri atas 1.000 data calon peminjam. Dataset memiliki lima atribut prediktor, yaitu Age, Gender, Income, LoanAmount, dan CreditScore, serta satu atribut target Loan\_Status yang mencakup dua kelas, yaitu Approved dan Rejected. Dataset ini dimanfaatkan sebagai masukan dalam proses pelatihan dan pengujian system klasifikasi menggunakan algoritma Random Forest.

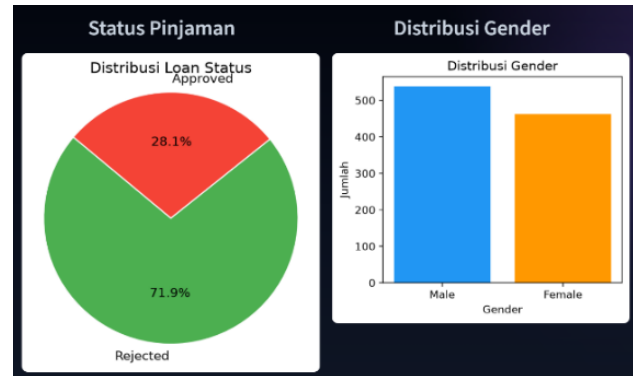
Preview Dataset

|    | applicant_id | Age | Gender | Income | LoanAmount | CreditScore | Loan_Status |
|----|--------------|-----|--------|--------|------------|-------------|-------------|
| 0  | 1            | 590 | Female | 197900 | 17800      | 6340        | Rejected    |
| 1  | 2            | 490 | Male   | 419110 | 109080     | 6260        | Rejected    |
| 2  | 3            | 350 | Male   | 644590 | 73030      | 4120        | Rejected    |
| 3  | 4            | 280 | Female | 393630 | 156860     | 7120        | Approved    |
| 4  | 5            | 410 | Male   | 350490 | 49830      | 4480        | Rejected    |
| 5  | 6            | 590 | Male   | 481805 | 105030     | 5170        | Rejected    |
| 6  | 7            | 390 | Female | 548250 | 179050     | 4530        | Rejected    |
| 7  | 8            | 430 | Male   | 310820 | 140700     | 5370        | Rejected    |
| 8  | 9            | 310 | Male   | 156610 | 132320     | 4520        | Rejected    |
| 9  | 10           | 310 | Female | 524120 | 55830      | 4430        | Rejected    |
| 10 | 11           | 440 | Female | 634790 | 69460      | 7020        | Approved    |

Gambar 3. Halaman Dataset

Gambar 3 menunjukkan halaman Dataset pada aplikasi yang digunakan untuk menampilkan data awal sebelum dilakukan proses *preprocessing*. Halaman ini menyajikan sebagian data beserta atribut yang digunakan sehingga pengguna dapat memverifikasi struktur dataset, tipe data, dan informasi setiap atribut sebelum memasuki tahap pengolahan data.

Selanjutnya dilakukan analisis terhadap distribusi kelas pada atribut Loan\_Status untuk mengetahui proporsi data pada masing-masing kelas.



Gambar 4. Distribusi Status Persetujuan Pinjaman

### C. Pelatihan Model Random Forest

Setelah melalui tahapan preprocessing dan transformation, dataset dibagi menjadi 80% data latih (training data) dan 20% data uji (testing data). Pembagian tersebut dilakukan untuk model mampu mengenali pola hubungan antara atribut Age, Gender, Income, LoanAmount, dan CreditScore terhadap status persetujuan pinjaman (Loan\_Status) pada data pelatihan, kemudian mengevaluasi kinerja model dalam mengklasifikasikan sampel yang belum pernah dipelajari pada tahap pelatihan.

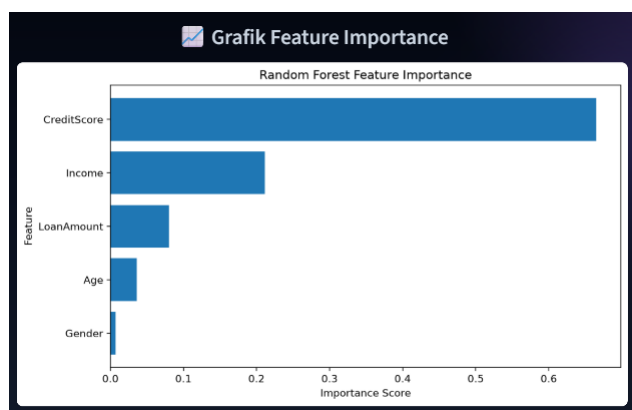
Proses pelatihan dilaksanakan menggunakan algoritma Random Forest Classifier yang menggunakan parameter *n\_estimators* sebesar 100 dan *random\_state* = 42. Selama proses pelatihan, algoritma membangun sejumlah Decision Tree berdasarkan data hasil bootstrap sampling. Masing – masing pohon keputusan memberi hasil prediksi secara independen, selanjutnya seluruh hasil prediksi dikombinasikan menggunakan mekanisme majority voting untuk menentukan keputusan akhir berupa status Approved atau Rejected.

Model yang sudah selesai melalui proses pelatihan selanjutnya digunakan dalam proses prediksi terhadap sampel pengujian. Output prediksi tersebut menjadi dasar dalam proses pengujian melalui Confusion Matrix, Accuracy,

Precision, Recall, dan F1-Score. Melalui proses tersebut, performa model dapat dinilai berdasarkan kemampuannya dalam mengklasifikasikan status persetujuan pinjaman secara akurat dan konsisten.

#### D. Analisis Nilai Feature Importance

Analisis Feature Importance dilakukan untuk mengevaluasi besarnya kontribusi setiap atribut terhadap proses klasifikasi status persetujuan pinjaman menggunakan algoritma Random Forest. Hasil analisis ini memberikan informasi mengenai fitur yang paling berpengaruh terhadap keputusan model, sehingga dapat digunakan sebagai dasar dalam memahami faktor-faktor yang memengaruhi persetujuan pinjaman.



Gambar 5. Distribusi Status Persetujuan Pinjaman

Tabel 3. Nilai Feature Importance

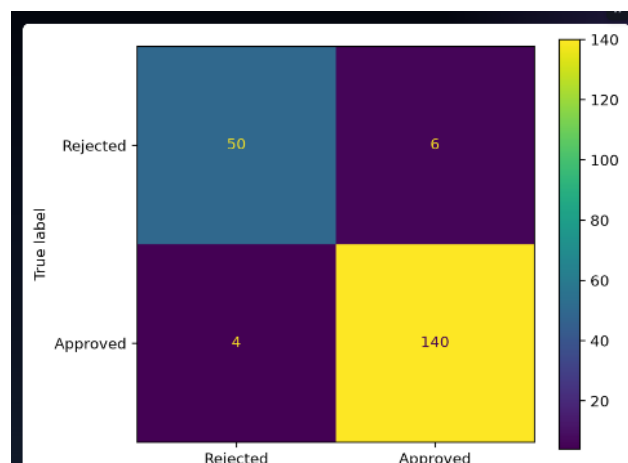
| Atribut     | Nilai Importance |
|-------------|------------------|
| CreditScore | 0,67             |
| Income      | 0,21             |
| LoanAmount  | 0,08             |
| Age         | 0,04             |
| Gender      | 0,01             |

Berdasarkan Gambar 5 dan Tabel 3, atribut CreditScore memiliki nilai feature importance tertinggi, yaitu sekitar 0,67, sehingga menjadi faktor yang paling dominan dalam menentukan status persetujuan pinjaman. Selanjutnya, atribut Income berada pada urutan kedua dengan nilai 0,21, diikuti oleh LoanAmount sebesar 0,08. Hasil ini menunjukkan bahwa keputusan persetujuan pinjaman lebih banyak dipengaruhi oleh kondisi finansial dan kemampuan calon

peminjam dalam memenuhi kewajiban pembayaran pinjaman.

Sementara itu, atribut Age dan Gender memiliki nilai feature importance yang tergolong rendah, yaitu sebesar 0,04 dan 0,01 secara berurutan, sehingga memberikan kontribusi yang lebih kecil terhadap proses klasifikasi. Hal ini menunjukkan bahwa model lebih memprioritaskan atribut yang berkaitan dengan kemampuan finansial dan risiko kredit dibandingkan karakteristik demografis pemohon. Secara keseluruhan, hasil analisis Feature Importance mengindikasikan bahwa CreditScore serta Income menjadi faktor utama yang memengaruhi keputusan persetujuan pinjaman. Temuan ini mengindikasikan bahwa model Random Forest dapat mengidentifikasi atribut-atribut sehingga paling relevan pada proses klasifikasi, sehingga hasil yang diperoleh dapat menjadi dasar dalam mendukung pengambilan keputusan persetujuan pinjaman secara lebih objektif.[1]

#### E. Analisis Confusion Matrix



Gambar 6. Confusion Matrix Algoritma Random Forest

Confusion Matrix digunakan untuk mengevaluasi kemampuan algoritma Random Forest dalam melakukan klasifikasi status persetujuan pinjaman menggunakan data pengujian. Hasil pengujian mengindikasikan bahwa model mampu mengklasifikasikan 50 data Rejected dan 140 data Approved dengan benar. Selain itu, terdapat 6 data Rejected yang diklasifikasikan sebagai Approved (False

Positive), sedangkan 4 data Approved yang diklasifikasikan sebagai Rejected (False Negative).

Berdasarkan hasil tersebut, jumlah kesalahan klasifikasi yang dihasilkan relatif kecil dibandingkan dengan jumlah prediksi yang benar. Hal ini mengindikasikan bahwa model dapat membedakan setiap kelas secara akurat serta menghasilkan prediksi sehingga konsisten pada sampel pengujian.

Hasil Confusion Matrix juga menunjukkan bahwa jumlah False Positive dan False Negative yang relatif rendah berkontribusi terhadap nilai Accuracy, Precision, Recall, dan F1-Score yang tinggi. Oleh karena itu, model klasifikasi Random Forest dinilai menunjukkan kemampuan klasifikasi yang baik dan layak digunakan untuk memprediksi status persetujuan pinjaman.

#### F. Analisis Evaluasi Model

Evaluasi model dilakukan untuk mengevaluasi tingkat performa algoritma Random Forest dalam melakukan klasifikasi status persetujuan pinjaman menggunakan sampel pengujian. Pengukuran performa dilakukan berdasarkan hasil Confusion Matrix melalui metrik Accuracy, Precision, Recall, dan F1-Score. Metrik tersebut digunakan untuk mengetahui performa model dalam menghasilkan keluaran model yang akurat serta konsisten pada kedua kelas, yaitu Approved dan Rejected.

Tabel 4. Hasil Evaluasi Model

| Metrik    | Nilai  |
|-----------|--------|
| Accuracy  | 95,00% |
| Precision | 95,89% |
| Recall    | 97,22% |
| F1-Score  | 96,55% |

Berdasarkan hasil Confusion Matrix, diperoleh nilai True Positive (TP) = 140, True Negative (TN) = 50, False Positive (FP) = 6, dan False Negative (FN) = 4. Nilai-nilai tersebut digunakan untuk menghitung metrik evaluasi sebagai berikut.

#### a. Perhitungan Accuracy

$$\begin{aligned}\text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{140 + 50}{140 + 50 + 6 + 4} \\ &= \frac{190}{200} = 0,95 = 95,00\%\end{aligned}$$

#### b. Perhitungan Precision

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP + FP} \\ &= \frac{140}{140 + 6} \\ &= \frac{140}{146} = 0,9589 = 95,89\%\end{aligned}$$

#### c. Perhitungan Recall

$$\begin{aligned}\text{recall} &= \frac{TP}{TP + FN} \\ &= \frac{140}{140 + 4} \\ &= \frac{140}{144} = 0,9722 = 97,22\%\end{aligned}$$

#### d. Perhitunagn F1-Score

$$\begin{aligned}\text{F1 - Score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= \frac{2 \times 0,9589 \times 0,9722}{0,9589 + 0,9722} \\ &= 0,9655 = 96,55\%\end{aligned}$$

Berdasarkan hasil evaluasi, algoritma Random Forest berhasil memperoleh Accuracy sebesar 95,00%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Nilai Precision sebesar 95,89% menunjukkan bahwa prediksi pada kelas Approved memiliki tingkat ketepatan yang tinggi. Selain itu, nilai Recall sebesar 97,22% menunjukkan bahwa model dapat mengenali hampir seluruh sampel yang benar-benar termasuk dalam kategori Approved, sedangkan F1-Score sebesar 96,55% menunjukkan keseimbangan yang sangat baik antara nilai Precision dan Recall.[8] [12],

Secara keseluruhan, hasil evaluasi tersebut membuktikan bahwa algoritma Random Forest memiliki performa klasifikasi yang optimal dalam memprediksi status persetujuan pinjaman.

Tingginya nilai pada seluruh metrik evaluasi mengindikasikan bahwa model mampu memberikan hasil prediksi yang akurat, konsisten, serta memiliki kemampuan generalisasi yang baik terhadap data pengujian. Oleh karena itu, model yang dibangun layak digunakan sebagai pendukung pengambilan keputusan dalam proses analisis persetujuan pinjaman.

#### G. Analisis Hasil

Berdasarkan hasil evaluasi, algoritma Random Forest mampu mengklasifikasikan status persetujuan pinjaman dengan performa yang sangat baik, ditunjukkan oleh nilai Accuracy sebesar 95,00%, Precision sebesar 95,89%, Recall sebesar 97,22%, dan F1-Score sebesar 96,55%. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang tinggi dalam membedakan data pinjaman yang disetujui (Approved) maupun ditolak (Rejected).

Hasil analisis Feature Importance menunjukkan bahwa atribut CreditScore merupakan faktor yang paling berpengaruh dalam proses klasifikasi, diikuti oleh Income dan LoanAmount. Hal ini mengindikasikan bahwa keputusan persetujuan pinjaman lebih banyak dipengaruhi oleh kondisi finansial dan tingkat risiko kredit calon peminjam dibandingkan karakteristik demografis seperti Age dan Gender.

Secara umum, temuan penelitian ini memperlihatkan bahwa algoritma Random Forest layak diterapkan dalam pembangunan model klasifikasi persetujuan pinjaman. Selain menghasilkan tingkat akurasi yang tinggi, model juga mampu memberikan informasi mengenai faktor-faktor yang paling berpengaruh terhadap keputusan pinjaman sehingga dapat dimanfaatkan sebagai pendukung pengambilan keputusan pada lembaga keuangan.

#### IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Random Forest untuk mengklasifikasikan status persetujuan pinjaman menggunakan Loan Approval Dataset berdasarkan atribut Age, Gender,

Income, LoanAmount, dan CreditScore. Hasil pengujian menunjukkan bahwa model menunjukkan kemampuan untuk memberi kinerja klasifikasi yang sangat baik dengan capaian Accuracy mencapai 95,00%, Precision sebesar 95,89%, Recall sebesar 97,22%, dan F1-Score sebesar 96,55%. Hasil tersebut menunjukkan bahwa algoritma Random Forest mampu mengklasifikasikan status persetujuan pinjaman secara akurat dan konsisten.

Berdasarkan analisis Feature Importance, atribut CreditScore menjadi faktor yang paling berpengaruh dalam menentukan persetujuan pinjaman, diikuti oleh Income dan LoanAmount. Temuan ini menunjukkan bahwa kondisi finansial dan kualitas kredit calon peminjam memiliki peranan yang lebih dominan dibandingkan karakteristik demografis seperti Age dan Gender. Selain itu, model yang telah dibangun berhasil diimplementasikan ke dalam aplikasi berbasis Streamlit, sehingga dapat dimanfaatkan sebagai sistem pendukung keputusan untuk membantu proses analisis persetujuan pinjaman secara lebih cepat, objektif, dan efisien.

Penelitian berikutnya disarankan memanfaatkan dataset dengan jumlah sampel dan atribut yang lebih baik serta melakukan perbandingan performa Random Forest dengan algoritma klasifikasi lainnya atau menerapkan teknik optimasi parameter (hyperparameter tuning) untuk memperoleh model dengan tingkat akurasi yang lebih optimal.

#### REFERENSI

- [1] Andi and Thamrin, "Komparasi kinerja algoritma random forest dan support vector machine berbasis adaptive boosting untuk analisis kelayakan kredit nasabah," *J. Inf. Syst. Res.*, vol. XIII, no. 2, pp. 326–333, 2024.
- [2] T. I. R. I. Handayani, "Implementasi random forest untuk klasifikasi kelayakan kredit nasabah dalam pengelolaan risiko keuangan di PT BPR BKK Wonogiri," *J. Inf. Technol. Comput. Sci.*, vol. 8, no. 6, pp. 1937–1946, 2024.
- [3] A. Hafidh Ismu Azam and Qoiriah, "Prediksi kelayakan pinjaman berdasarkan profil risiko nasabah menggunakan logistic regression dan random forest," *J. Informatics Comput. Sci.*, vol. 7, pp. 301–312, 2025.
- [4] M. F. R. Luthfi Hadi Pratama, "Prediksi kelayakan pengajuan pinjaman di koperasi menggunakan random forest," *Tek. Inform.*, vol. 5, pp. 656–663, 2026.

- [5] [J. Nasional and S. Informasi, "Analisa prediksi kelayakan pemberian kredit pinjaman dengan metode random forest," *J. Nas. Teknol. dan Sist. Inf.*, vol. 2, pp. 79–89, 2021.
- [6] G. Surono and N. N. Pusparini, "Journal of technology information," *J. Technol. Inf.*, vol. 5, no. 2, pp. 99–104, 2020.
- [7] N. Hazizah, Sharipuddin, and A. Feranika, "Implementasi algoritma random forest dalam klasifikasi risiko gagal bayar kartu kredit pada nasabah bank," *J. Manaj. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 1050–1059, 2025, doi: 10.33998/jms.2025.5.1.2384.
- [8] P. A. Dwi Sandhi Bhakti and Agung Prasetyo, "Implementation of hyperparameter tuning in random forest algorithm for loan approval prediction," *J. Tek. Inform.*, vol. 5, no. 4, pp. 63–69, 2024, doi: 10.52436/1.jutif.2024.5.4.2032.
- [9] I. Sugiyarto, "Perbandingan kinerja algoritma data mining prediksi persetujuan kartu kredit," *Fakt. Exacta*, vol. 12, no. 3, p. 180, 2019, doi: 10.30998/faktorexacta.v12i3.4310.
- [10] S. Riyadi et al., "Perbandingan algoritma naive bayes dan random forest untuk klasifikasi kelayakan kredit pada PT BPR Magga Jaya Utama," *Proc. Technol. Eng. Comput.*, vol. 6, pp. 167–186, 2025.
- [11] D. P. Anindita Puspa Ayu Prayogi and Altha Inas Shofyana, "Prediksi credit card approval menggunakan algoritma random forest," *J. Ilmu Komput. dan Multimed.*, vol. 2, no. 1, pp. 21–25, 2025.
- [12] N. S. Moh. Ichlasul Amal Rois and Gifthera Dwilestari, "Prediksi persetujuan pinjaman menggunakan dataset loan approval menggunakan algoritma klasifikasi," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 1, pp. 1342–1347, 2025, doi: 10.36040/jati.v9i1.12486.
- [13] H. Putra and R. Rumini, "Comparative study of logistic regression, random forest, and XGBoost for bank loan approval classification," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2822–2835, 2025, doi: 10.30871/jaic.v9i5.10862.
- [14] M. F. Azis and N. D. Rahmawati, "Tinjauan hukum terhadap perjanjian pinjaman online dan penggunaan data konsumen aplikasi 'Kredit Pintar'," *Fortioro Law J.*, vol. 1, no. 1, p. 5, 2021.
- [15] Y. H. Omar Pahlevi, Amrin, "Implementasi algoritma klasifikasi random forest untuk penilaian kelayakan kredit," *J. Infortech*, vol. 5, no. 1, pp. 71–76, 2023, doi: 10.31294/infortech.v5i1.15829.