

Implementasi Algoritma Random Forest Dalam Klasifikasi Risiko Gagal Bayar Kartu Kredit Pada Nasabah Bank

Nur Hazizah¹, Sharipuddin², Ayu Feranika³

^{1,2}Fakultas Ilmu Komputer, ³Fakultas Ilmu Manajemen dan Bisnis, Universitas Dinamika Bangsa, Kota Jambi, Indonesia

Jl. Jendral Soedirman Thehok-Jambi, telp (0741) 35095

Email: ¹zizahnurhazizah67@gmail.com, ²sharifbuhaira@gmail.com, ³ayuferanika2@gmail.com

Email Penulis Korespondensi: ayuferanika2@gmail.com

Submitted :
04 Januari 2025

Revision :
23 Januari 2025

Accepted:
23 Maret 2025

Published:
28 Maret 2025

Abstrak— Situasi keuangan klien yang tidak stabil atau manajemen keuangan yang buruk merupakan penyebab umum risiko gagal bayar kartu kredit, yang merupakan masalah besar bagi industri perbankan. Tujuan dari penelitian ini adalah untuk menggunakan algoritma *Random Forest* untuk mengkategorikan risiko gagal bayar kartu kredit klien bank. Metode yang digunakan mencakup pengumpulan data, *preprocessing*, pelatihan model, evaluasi, dan analisis faktor penting menggunakan dataset publik yang berisi 30.000 data. Temuan penelitian menunjukkan bahwa algoritma *Random Forest* mencapai tingkat akurasi 81% dalam hal penilaian risiko gagal bayar. Algoritma tersebut secara signifikan dipengaruhi oleh kriteria utama seperti riwayat pembayaran sebelumnya, batas kredit, dan jumlah tagihan. Meskipun model menunjukkan performa yang baik untuk nasabah yang tidak gagal bayar, tantangan tetap ada pada klasifikasi nasabah yang gagal bayar, terutama karena ketidakseimbangan data. Selain meletakkan dasar untuk perbaikan model di masa mendatang, penelitian ini menunjukkan bahwa metode *Random Forest* berguna untuk membantu pengambilan keputusan sektor perbankan.

Kata Kunci: Random Forest; Klasifikasi Risiko Kredit; Gagal Bayar; Kartu Kredit; Nasabah Bank; Machine Learning

Abstract— Credit card default risk is a major challenge in the banking industry, often caused by unstable financial conditions or poor financial management by customers. This study aims to classify the credit card default risk of bank customers by implementing the Random Forest algorithm. The methodology includes data collection, preprocessing, model training, evaluation, and key factor analysis using a public dataset containing 30,000 records. According to the findings, the Random Forest algorithm gets 81% accuracy when classifying default risk, with important criteria including past payment status, credit limit, and billing amount playing a significant role. Classifying defaulting consumers remains challenging, especially due to data asymmetry, however the model does a good job of identifying customers who do not default. The results of this study show that the Random Forest algorithm is useful for banking sector decision-making and lay the groundwork for future model improvements.

Keywords: Random Forest; Credit Risk Classification; Default; Credit Card; Bank Customers; Machine Learning

1. PENDAHULUAN

Kartu kredit adalah instrumen pembayaran plastik yang diterbitkan oleh bank tertentu. Kartu tersebut dapat digunakan untuk melakukan pembelian secara online atau di toko fisik menggunakan kredit pemegang kartu atau pemegang kartu yang disebutkan namanya, dan juga dapat digunakan untuk menarik uang tunai dari ATM hingga batas yang ditetapkan bank [1]. Kartu kredit merupakan salah satu cara bagi pembeli yang bertanggung jawab untuk membuktikan identitas dan kemampuan mereka dalam membayar [2]. Kartu kredit merupakan salah satu cara untuk membayar barang dan jasa secara daring atau di toko dengan menggunakan layanan yang ditawarkan oleh lembaga keuangan, atau untuk mendapatkan uang tunai dari lembaga yang sama [3].

Penjual kartu kredit dari berbagai bank dapat terlihat di berbagai pusat perbelanjaan, dan mereka semua bersaing untuk mendapatkan pelanggan. Pemegang kartu kredit dibujuk untuk memiliki hak istimewa kartu kredit dengan berbagai penawaran kemudahan dan fasilitas, beserta hadiah. Tujuan utama kartu kredit adalah untuk memfasilitasi pembayaran; kartu kredit tidak menghasilkan pendapatan tambahan. Kemudahan penggunaan merupakan keuntungan utama penggunaan kartu kredit yang kita miliki. Terlepas dari semua manfaatnya, sebagian orang menggunakan kartu kredit mereka terlalu sering. Oleh karena itu, mereka terlilit utang yang sangat besar dan tidak mampu melunasi pokok utang beserta bunganya. Akibatnya, gagal bayar dapat terjadi sebagai akibat dari penggunaan kartu kredit yang berlebihan dan ketidakmampuan untuk menahan berbagai godaan [4].

Ketika pemegang kartu gagal melakukan sebagian atau seluruh pembayarannya sebagaimana disepakati dalam perjanjian, hal ini dikenal sebagai gagal bayar atau kredit macet. Nasabah umumnya gagal bayar kartu kredit karena alasan-alasan berikut: untuk meningkatkan rasa harga diri, merasa berharga, mengalihkan pendapatan, meningkatkan kebahagiaan, memenuhi kebutuhan, dan menghindari pengeluaran melebihi kemampuan. Utang kartu kredit merupakan masalah yang umum terjadi karena membuat pengeluaran masyarakat melebihi pendapatan dan terus meningkat, sehingga memaksa mereka bekerja keras untuk melunasi utang [5].

Penelitian lain yang terkait dilakukan oleh Prasojo dan Haryatmi, 2021 [6] Pada kegiatan penelitian ini menggunakan klasifikasi algoritma Random Forest, Penelitian dilakukan untuk melihat implementasi metode klasifikasi dengan algoritma Random Forest pada german data credit. Hasil dari tahapan pengujian akhir diperoleh nilai performa dengan menggunakan AUC mendapatkan nilai sebesar 0,83 (83%) dan termasuk ke dalam *Very Good Model*. penelitian selanjutnya dilakukan oleh Ibnu dan Rachmatullah, 2023 [7] untuk membantu mengklasifikasi kegagalan pembayaran kredit nasabah bank salah satunya yaitu yang dilakukan oleh dengan membandingkan empat metode klasifikasi yaitu : Random Forest, KNN, SVM, dan MLP, dengan penerapan SMOTE pada dataset menghasilkan peningkatan kinerja. Untuk kategorisasi debitur berdasarkan kualitas kredit, penelitian lain yang dilakukan oleh Rochdi Wasono, 2022 [8] menemukan bahwa Naive Bayes mencapai tingkat akurasi 95,33% dan Random Forest mencapai tingkat akurasi 98,16%.

Zailani dan Hanun, 2020 [9] menemukan bahwa data kredit bermasalah Koperasi Mitra Sejahtera memiliki tingkat akurasi 87,88% ketika dianalisis menggunakan metode random forest. Model prediksi penerimaan aplikasi pinjaman yang sangat akurat diperoleh pada penelitian sebelumnya oleh Yoga, Agung, dan Wahyu, 2021 [10] menggunakan metode random forest yang dipadukan dengan bagging dan optimasi GA. Penelitian ini menemukan bahwa metode random forest mencapai AUC (+) sebesar 0,500 dan AUC (-) sebesar 0,000 ketika digunakan untuk mengklasifikasikan dataset pemasaran bank, sehingga menghasilkan akurasi sebesar 88,30%. Mengacu pada penelitian mereka pada tahun 2021 [11], Religia, Nugroho, dan Hadikristanto melaporkan bahwa hasil pengujian menghasilkan kinerja klasifikasi dataset pemasaran bank ketika menggunakan algoritma *Random Forest* untuk tujuan mengklasifikasikan prediksi penerimaan aplikasi pinjaman. Area di bawah kurva (AUC) sebesar 0,500 dan akurasi sebesar 88,30% dihasilkan dengan menggunakan pendekatan random forest yang dipilih.

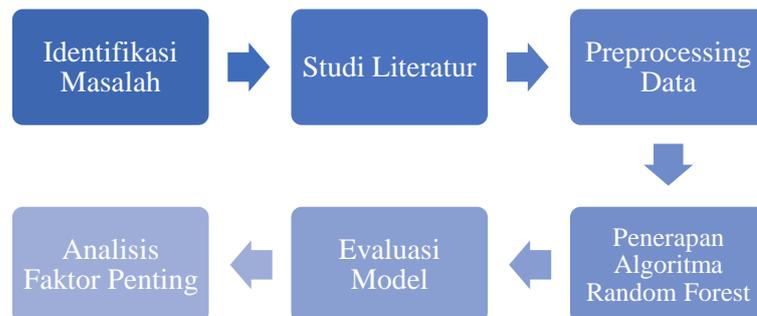
Teknik agregasi pohon keputusan digunakan dalam hutan acak, salah satu metode berbasis regresi dan klasifikasi. Peneliti memilih pendekatan ini karena secara efektif menangani data yang tidak lengkap, menangani volume besar data pelatihan, memiliki tingkat kesalahan yang rendah, dan memberikan akurasi klasifikasi yang baik [12]. Untuk menghasilkan prediksi yang lebih andal dan tepat, metode *Random Forest* dapat menggabungkan output dari banyak pohon keputusan. Di antara manfaat utama metode ini adalah kapasitasnya untuk menangani dataset besar, variabel berkorelasi, dan mencegah overfitting [13]. Di antara banyak pendekatan untuk regresi dan klasifikasi, hutan acak adalah pilihan yang populer. Teknik ini merupakan kumpulan algoritma pembelajaran yang disusun dan digabungkan dengan memanfaatkan pohon keputusan sebagai dasar pengklasifikasi [14]. Dalam hal pembelajaran mesin, *Random forest* merupakan salah satu algoritma yang paling terkenal dan efektif. Teknik ini termasuk dalam kategori pembelajaran ensemble dan menggunakan beberapa model sederhana untuk membuat prediksi yang lebih baik secara keseluruhan [15]. Dengan menggabungkan bagging dengan pemilihan atribut acak, *Random forest* dapat dibangun. Pohon keputusan dilatih menggunakan teknik *Classification and Regulation Tree* (CART). Tidak akan ada pemangkasan pohon keputusan setelah mencapai ukuran maksimumnya. Akibatnya, hutan lahir dari kelompok pohon ini [16].

Berdasarkan justifikasi ini, penulis akan menyelidiki metode untuk mengidentifikasi dan mengkategorikan risiko kredit nasabah menggunakan pemodelan Machine Learning. Maka “Implementasi Algoritma *Random Forest* Dalam Klasifikasi Risiko Gagal Bayar Kartu Kredit Pada Nasabah Bank” adalah judul kerja karya penelitian penulis, yang menggunakan teknik *Random Forest* Machine Learning.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Untuk mengatasi masalah yang diangkat, kerangka kerja penelitian ini akan diikuti. Gambar 2.1 menampilkan kerangka kerja penelitian yang digunakan:



Gambar 2.1 Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian pada gambar 2.1, maka dapat dijabarkan pembahasan dari masing-masing tahap penelitian tersebut sebagai berikut:

1. Identifikasi Masalah

Pada tahap ini, kerangka kerja digunakan untuk memilih topik penelitian dengan menganalisis fenomena terkini. Menggunakan algoritma *Random Forest* untuk mengkategorikan risiko gagal bayar kartu kredit bagi nasabah bank merupakan tantangan yang disorot.

2. Studi Literatur

Pengumpulan informasi dari sumber yang memiliki reputasi baik (jurnal, makalah, dll.) dilakukan pada tahap ini dengan mempertimbangkan landasan teori. Sumber-sumber tersebut juga didapatkan dari berbagai referensi, baik yang tertulis maupun yang berbasis media elektronik.

3. Pengumpulan Data

Tahap pertama adalah pengumpulan data yang relevan dengan klasifikasi risiko kredit, yang dapat berasal dari sumber seperti Kaggle. Contohnya, dataset "*Credit card*" yang mencakup variabel seperti usia, pekerjaan, jumlah pinjaman, dan riwayat kredit. Data yang dikumpulkan akan diunduh, disimpan dalam format yang sesuai, dan divalidasi untuk memastikan kelengkapan dan kesesuaian dengan penelitian sebelum digunakan dalam analisis lebih lanjut.

4. Preprocessing Data

Saat menangani masalah data seperti nilai yang hilang, outlier, atau variabel kategoris, prapemrosesan data merupakan langkah penting dalam pembelajaran mesin [17]. Langkah-langkah yang dilakukan meliputi:

a. *Handling Missing Values*

Mengidentifikasi dan menangani data yang hilang dengan metode seperti penghapusan baris atau *imputation* (mengganti dengan rata-rata, median, atau mode).

b. *One-Hot Encoding*

Mengubah variabel kategorikal menjadi format biner agar dapat digunakan oleh algoritma *machine learning*.

c. Normalisasi/Standardisasi

Menyamakan skala fitur agar model bekerja lebih baik, meskipun algoritma seperti *Random Forest* tidak terlalu sensitif terhadap skala data.

5. Penerapan Algoritma *Random Forest*

Setelah data diproses, model *Random Forest* dilatih dengan data yang sudah dibersihkan. Langkah-langkahnya meliputi:

a. Pemilihan *Hyperparameters*

Menemukan nilai optimal untuk hiperparameter seperti *n_estimators*, (*max_depth*), dan fitur *max_features* untuk mengoptimalkan kinerja.

b. Pelatihan Model

Membagi data menjadi *training* set dan *testing* set, kemudian melatih model dengan *training* set menggunakan *Random Forest*.

c. *Out-Of-Bag (OOB) Error*

Salah satu cara untuk menilai generalisasi model adalah dengan menggunakan kesalahan out-of-bounds (OOB) selama pelatihan. Metrik ini menilai seberapa baik model bekerja pada data yang tidak digunakan untuk pelatihan.

6. Evaluasi Model

Metrik berikut digunakan untuk evaluasi setelah model dilatih:

a. Akurasi

Menilai keakuratan prakiraan; namun, ia kehilangan sebagian kredibilitasnya saat berhadapan dengan data yang tidak seimbang.

b. *Precision*

Rasio prediksi positif yang benar, mengukur ketepatan model dalam mengidentifikasi nasabah berisiko tinggi.

c. *Recall*

Mengukur seberapa banyak nasabah berisiko tinggi yang terdeteksi, penting untuk meminimalkan risiko kredit macet.

d. *F1-score*

Nilai yang menggabungkan *precision* dan *recall*, memberikan keseimbangan jika terdapat ketidakseimbangan kelas.

7. Analisis Faktor Penting

Setelah evaluasi model, langkah selanjutnya adalah menganalisis fitur penting (*feature importance*) menggunakan *Random Forest*. Fitur penting dihitung berdasarkan seberapa sering variabel digunakan untuk membagi data dan kontribusinya terhadap akurasi model. Analisis ini membantu mengidentifikasi faktor-faktor utama, seperti pendapatan nasabah, yang mempengaruhi keputusan kredit, sehingga lembaga keuangan dapat memperbaiki proses evaluasi kredit.

2.2 Alat Bantu Penelitian

Dalam penerepan Implementasi Algoritma *Random Forest* Dalam Klasifikasi Risiko Gagal Bayar Kartu Kredit Pada Nasabah Bank dibutuhkan beberapa *hardware* dalam dan *software* dalam pembuatannya, yaitu :

1. Perangkat Keras (*Hardware*)
 - a. Laptop *Processor* Intel i3-HP 240 G7
 - b. RAM 8 GB
 - c. SSD 128 GB
 - d. HDD 1000 TB
2. Perangkat lunak (*Software*)
 - a. Sistem Operasi Windows 10
 - b. Google *Colab*
 - c. Kaggle
 - d. *Note Pad*
 - e. Google Chrome
 - f. Serta perangkat lunak lain yang mendukung penelitian

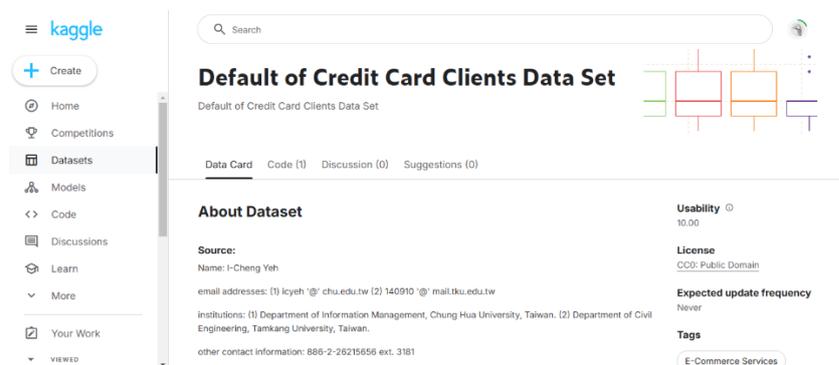
3. HASIL DAN PEMBAHASAN

3.1 Gambaran Umum

Salah satu masalah terbesar yang dihadapi bank adalah bahaya gagal bayar kartu kredit. Sejumlah variabel, termasuk situasi keuangan konsumen yang genting, pergeseran ekonomi, atau pengelolaan uang yang buruk, berkontribusi terhadap kegagalan mereka untuk memenuhi kewajiban pembayaran. Risiko ini dapat memberikan dampak negatif yang signifikan, baik pada institusi perbankan dalam bentuk kerugian finansial, maupun pada nasabah yang mengalami kesulitan dalam mengelola utangnya. Ini juga merupakan komponen penting yang harus dipertimbangkan dalam manajemen kredit karena semakin kompleksnya faktor-faktor yang memengaruhi risiko ini.

3.2 Pengumpulan Dan Persiapan Data

Penelitian ini menggunakan dataset “*default of credit card clients*” yang berisi informasi terkait risiko gagal bayar nasabah. Pada Kolom *dpnm* digunakan sebagai label target untuk mengklasifikasikan status risiko gagal bayar nasabah.



Gambar 3.1 Tampilan *default of credit card clients* di kaggle

Data klien kartu kredit default dari *Kaggle* ditampilkan pada Gambar 3.1. *Kaggle* adalah platform online yang menyediakan berbagai dataset, alat untuk analisis data, serta ruang kolaborasi bagi data *scientist*, analis data, dan pengembang. Salah satu fitur unggulannya, *Kaggle Dataset*, memungkinkan pengguna untuk menemukan, membagikan, dan memanfaatkan dataset dari berbagai sumber. Dataset ini merupakan salah satu dataset yang tersedia di *Kaggle* dan digunakan dalam penelitian ini. Dataset dapat diakses melalui link : <https://www.kaggle.com/datasets/tunguz/default-of-credit-card-clients-data-set>. Pada dataset ini memiliki 30.000 data dan memiliki 25 atribut antara lainnya, *Id*, *Limit_Bal*, *Gender*, *Education*, *Marrige*, *Age*, *Pay_1*, *Pay_2*, *Pay_3*, *Pay_4*, *Pay_5*, *Pay_6*, *Bill_Amt1*, *Bill_Amt2*, *Bill_Amt3*, *Bill_Amt4*, *Bill_Amt5*, *Bill_Amt6*, *Pay_Amt1*,

Pay_Amt2, Pay_Amt3, Pay_Amt4, Pay_Amt5, Pay_Amt6, dan *dpm*. Dari kolom *dpm* memiliki potensi dalam menentukan 'gagal bayar' dan 'tidak gagal bayar' pada nasabah kartu kredit.

3.3 Pengumpulan Data dan Pemahaman Data

Representasi data adalah langkah awal yang penting dalam analisis dataset. Meskipun menjamin kualitas data yang digunakan untuk membangun model prediksi, representasi data memungkinkan kita melihat tren yang mungkin mengarah pada penipuan. Dalam penelitian ini, penulis menggunakan dataset yang berisi 30.000 data dengan 25 atribut yang menjadi fokus analisis. Seluruh atribut tersebut digunakan sebagai variabel utama dalam penelitian ini. yang mana seluruh atributnya yaitu:

ID: Nomor unik untuk mengidentifikasi setiap nasabah.

1. *Limit_BAL*: Batas kredit total yang diberikan kepada.
2. *Gender*: Jenis kelamin nasabah.
3. *Education*: Tingkat pendidikan nasabah.
4. *Marriage*: Status hubungan konsumen.
5. *Age*: Usia pelanggan dalam tahun.
6. *Pay_1*: Kemajuan penggantian biaya bulan pertama.
7. *Pay_2*: Pembaruan situasi pembayaran bulan kedua.
8. *Pay_3*: Pembaruan status pembayaran bulan ketiga.
9. *Pay_4*: Kemajuan pembayaran bulan keempat.
10. *Pay_5*: Laporan kemajuan menuju tanggal pembayaran di bulan kelima.
11. *Pay_6*: Pembaruan status pembayaran bulan keenam.
12. *Bill_AMT1*: Jumlah yang harus dibayar dalam 30 hari pertama.
13. *Bill_AMT2*: Biaya untuk bulan kedua secara keseluruhan.
14. *Bill_AMT3*: Jumlah tagihan pada bulan ketiga.
15. *Bill_AMT4*: Jumlah total yang harus dibayar di bulan keempat.
16. *Bill_AMT5*: Tagihan terakhir yang harus dibayar di bulan kelima.
17. *Bill_AMT6*: Terakhir, total tagihan untuk bulan keenam.
18. *Pay_AMT1*: Pembayaran penuh diterima dalam 30 hari pertama.
19. *Pay_AMT2*: Pembayaran selama dua bulan selesai.
20. *Pay_AMT3*: Total pembayaran bulan ketiga diterima.
21. *Pay_AMT4*: Bulan keempat melihat jumlah total yang dibayarkan.
22. *Pay_AMT5*: Pembayaran dilakukan secara penuh selama bulan keenam.
23. *Pay_AMT6*: Dana yang dicairkan selama bulan keenam
25. *dpm*: Status pembayaran akhir nasabah (apakah bermasalah atau tidak).

3.4 Perhitungan *Random Forest* Pada Dataset Risiko Kartu Kredit Menggunakan *Google Colab*

a. Memuat dan Menampilkan Data

Pada bagian ini, peneliti akan memuat dan menampilkan data dari dataset *default of credit card clients*. Proses ini meliputi normalisasi fitur numerik serta perubahan fitur kategorikal menjadi format yang dapat diproses oleh model. Berikut ini langkah-langkah yang diambil:

1. Mengimport Library

Pada tahap ini, sebelum melakukan pemrosesan data, peneliti akan memulai dengan memasukkan perintah impor pustaka yang digunakan untuk menjelaskan fungsi dari setiap pustaka yang diimpor dalam proyek atau analisis. Penjelasan ini bertujuan untuk menggambarkan peran masing-masing pustaka dalam mendukung proses analisis atau pengolahan data. Berikut adalah tahapan yang dilakukan:

```
#Langkah 1: Import Library
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, classification_report
```

Gambar 3.2 Import library

Pada Gambar 3.2 di atas, pustaka-pustaka penting diimpor untuk analisis data dan pemodelan. Pustaka seperti *pandas* dan *numpy* digunakan untuk manipulasi data dan perhitungan numerik, sementara *seaborn* dan *matplotlib* digunakan untuk visualisasi data. Dengan *Scikit-learn* dapat membagi data ke dalam set pelatihan dan pengujian, menggunakan *LabelEncoder* untuk melakukan pra-proses label, dan menilai model dengan bantuan matriks kebingungan dan sejumlah metrik seperti recall, akurasi, presisi, dan F1-score. Selain itu, pustaka *imblearn* digunakan untuk menangani ketidakseimbangan kelas dengan teknik *SMOTE*, dan *RandomForestClassifier* digunakan untuk membangun model klasifikasi berbasis *Random Forest*.

2. Memasukkan file dataset

```
# Langkah 2: Load Dataset
file_path = '/content/default of credit card clients.csv' # Ganti dengan path dataset Anda
data = pd.read_csv(file_path, header=1)
```

Gambar 3.3 memasukkan file dataset csv

Pada Gambar 4.6, ditampilkan kode *Python* yang digunakan untuk memuat dataset dalam format file CSV ke dalam program. Lokasi file CSV yang bernama "*default of credit card clients.csv*" disimpan dalam variabel *file_path*, yang menunjukkan bahwa file tersebut berada di direktori */content/*. Data dibaca dari file CSV dan disimpan ke dalam *DataFrame* yang disebut data menggunakan metode *pd.read_csv(file_path)* dari pustaka *Pandas*. Dengan menggunakan kode ini, data yang telah dimuat ke dalam *DataFrame* tersebut dapat diproses lebih lanjut menggunakan berbagai fungsi yang tersedia dalam pustaka *Pandas*.

3. Pra-pemrosesan Data

```
#Pra-pemrosesan data
# mendeteksi kolom 'dpmn' sebagai target (gagal bayar)
# dan kolom lain sebagai fitur
# Gantilah 'dpmn' dengan nama kolom target

# Misal: target kolom 'dpmn' dan fitur yang lain
X = df.drop('dpmn', axis=1) # Fitur
y = df['dpmn']             # Target
```

Gambar 3.4 Pra-pemrosesan data

Gambar 3.4 menunjukkan langkah pra-pemrosesan data untuk memisahkan fitur dan target dalam dataset. Kolom *dpmn* digunakan sebagai target (label) untuk memprediksi gagal bayar, sementara kolom lainnya dianggap sebagai fitur. Baris *X = df.drop('dpmn', axis=1)* menghapus kolom *dpmn* dari *DataFrame* *df* dan menyimpan sisanya sebagai *DataFrame* *X*, yang berisi fitur. Baris *y = df['dpmn']* mengambil kolom *dpmn* sebagai variabel target dan menyimpannya dalam variabel *y*. Langkah ini merupakan persiapan awal sebelum melatih model *machine learning*.

4. Split Data/Bagi Data Train, Test

```
# Membagi data menjadi data train (70%) dan data test (30%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

Gambar 3.5 Split Data/Bagi Data Train, Test

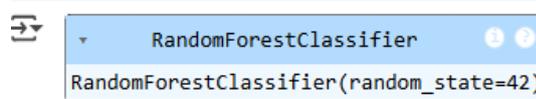
Gambar 3.5 menampilkan hasil penggunaan metode *train_test_split* untuk mempartisi set data ke dalam set pelatihan dan pengujian. Untuk melatih model, kami menggunakan variabel *X_train* dan *y_train*. Untuk menguji kinerjanya, kami menggunakan variabel *X_test* dan *y_test*. Dengan *test_size=0.3*, kami memberi tahu model untuk menggunakan 70% data untuk pelatihan dan 30% untuk pengujian. Selain itu, parameter *random_state=42* memastikan bahwa pembagian data dilakukan secara konsisten setiap kali kode dijalankan. Langkah ini penting untuk mengevaluasi performa model secara objektif dengan memisahkan data pelatihan dan pengujian.

5. Melatih Model *Random Forest*

```
# Membuat model Random Forest
rf_model = RandomForestClassifier(random_state=42)

# Melatih model menggunakan data latih
rf_model.fit(X_train, y_train)
```

Gambar 3.6 Melatih Model *Random Forest*



Gambar 3.7 Tampilan Model *Random Forest*

Model klasifikasi berbasis algoritma *Random Forest* ditampilkan pada Gambar 3.7. Model tersebut dibangun dan disimpan dalam variabel *rf_model* dengan menggunakan `RandomForestClassifier(random_state=42)`. Agar fitur *X_train* dan label *y_train* dapat digunakan sebagai data pelatihan dalam model, fungsi `fit(X_train, y_train)` digunakan. Tujuan model ini adalah untuk meningkatkan akurasi prediksi dengan mempelajari pola dalam data pelatihan dan kemudian menggabungkan hasil dari banyak pohon keputusan.

6. Grafik *Result*

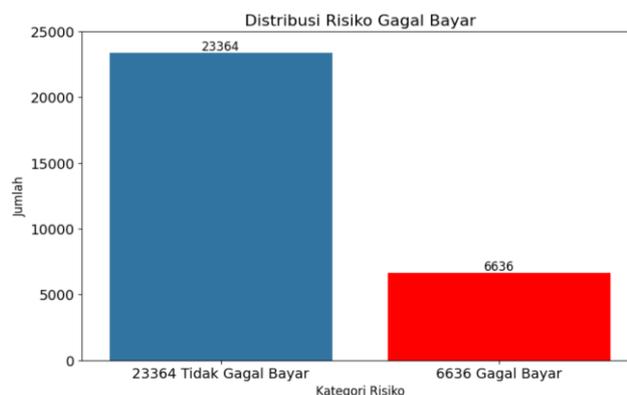
Untuk menemukan kemungkinan default atau non-default, analisis data dilakukan pada tahap ini. Proses ini dilakukan dengan memanfaatkan data yang telah melalui tahap pengolahan dan pemrosesan, menggunakan dataset *default of credit card clients*.

```
# Menghitung distribusi kategori pada kolom 'dpmn'
risk_distribution = data['dpmn'].value_counts()

# Menampilkan grafik
plt.show()
```

Gambar 3.8 Perintah Untuk Menampilkan Grafik Distribusi Risiko Gagal Bayar

Pada Gambar 3.8 dilakukan penghitungan distribusi kategori risiko gagal bayar pada kolom *dpmn* untuk melihat jumlah nasabah yang termasuk dalam kategori "Gagal Bayar" dan "Tidak Gagal Bayar." Distribusi ini dihitung menggunakan fungsi `value_counts()` pada kolom *dpmn*, yang menghasilkan jumlah masing-masing kategori. Grafik distribusi kemudian ditampilkan untuk memvisualisasikan data tersebut, di mana setiap batang pada grafik mewakili jumlah nasabah dalam masing-masing kategori, membantu memberikan gambaran yang lebih jelas mengenai proporsi risiko gagal bayar di dalam dataset.



Gambar 3.9 Grafik Distribusi Risiko Gagal Bayar

Gambar 3.9 menunjukkan distribusi risiko gagal bayar berdasarkan kategori. Dalam diagram batang ini, terdapat dua kategori utama: "23364 Tidak Gagal Bayar" yang diwakili oleh batang biru dan "6636 Gagal Bayar" yang diwakili oleh batang merah. Dari data ini, terlihat bahwa jumlah individu yang tidak gagal bayar (23,364) jauh lebih besar dibandingkan dengan mereka yang gagal bayar (6,636). Grafik ini memberikan gambaran jelas mengenai proporsi risiko gagal bayar, di mana mayoritas individu berada dalam kategori tidak gagal bayar.

3.5 Deteksi dengan Pengukuran Akurasi

Bagian ini menjelaskan proses penerapan dan evaluasi model menggunakan algoritma *Random Forest* untuk mengklasifikasikan risiko gagal bayar kartu kredit pada nasabah bank, serta menghitung tingkat akurasi model yang diperoleh.

```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import precision_recall_fscore_support, accuracy_score

# Membuat model Random Forest
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train) # Melatih model dengan data pelatihan

# Prediksi dengan data pengujian
y_pred = model.predict(X_test)

# Menghitung akurasi model
accuracy = accuracy_score(y_test, y_pred)
print(f"Akurasi Model Random Forest: {accuracy:.2f}")

# Confusion Matrix dengan Akurasi Ditampilkan
cm = confusion_matrix(y_test, y_pred)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=model.classes_)
disp.plot(cmap='Blues')
plt.title(f"Confusion Matrix (Akurasi: {accuracy:.2f})") # Menambahkan akurasi di judul
plt.show()

# Mendapatkan metrik dari classification report
precision, recall, f1, _ = precision_recall_fscore_support(y_test, y_pred, average=None, labels=[0, 1])

# Membuat bar chart untuk Precision, Recall, dan F1-score
metrics = ['Precision', 'Recall', 'F1-score']
class_0 = [precision[0], recall[0], f1[0]]
class_1 = [precision[1], recall[1], f1[1]]

x = range(len(metrics))

plt.figure(figsize=(10, 6))
plt.bar(x, class_0, width=0.4, label="Class 0 (Tidak Gagal Bayar)", color='lightblue', align='center')
plt.bar(x + 0.4, class_1, width=0.4, label="Class 1 (Gagal Bayar)", color='salmon', align='center')

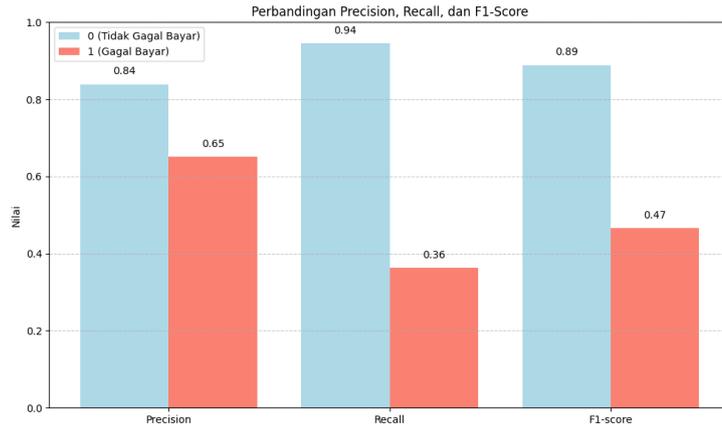
# Menambahkan detail ke grafik
plt.xticks([p + 0.2 for p in x], metrics)
plt.title("Perbandingan Precision, Recall, dan F1-Score")
plt.ylabel("Nilai")
plt.ylim(0, 1)
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

Gambar 3.10 Perintah Deteksi dengan Pengukuran Akurasi

Berikut Hasilnya:

🔄 Akurasi Model Random Forest: 0.82

Laporan Klasifikasi:	precision	recall	f1-score	support
0 (Tidak Gagal Bayar)	0.84	0.94	0.89	7009
1 (Gagal Bayar)	0.65	0.36	0.47	1991
accuracy			0.82	9000
macro avg	0.74	0.65	0.68	9000
weighted avg	0.80	0.82	0.80	9000



Gambar 3.11 Hasil Deteksi dengan Pengukuran Akurasi Dalam Bentuk Diagram

Pada gambar 3.11 Menunjukkan Model *Random Forest* memiliki akurasi sebesar 0.82, yang berarti bahwa model berhasil memprediksi dengan benar 82% dari data pengujian. Kelas 0 (klien non-default) dalam laporan klasifikasi menunjukkan skor f1 sebesar 0,89, recall sebesar 0,94, dan presisi sebesar 0,84 untuk model tersebut. Ini menunjukkan bahwa model tersebut memiliki recall yang tinggi dan presisi yang agak rendah dalam hal meramalkan klien non-default. Akurasi model adalah 0,65, recall adalah 0,36, dan f1-score adalah 0,47 untuk kelas 1, yang mewakili konsumen yang gagal bayar. Ini menunjukkan bahwa model tersebut kesulitan untuk memperkirakan konsumen yang gagal bayar karena recall-nya yang relatif buruk. Selain itu, estimasi makro dan rata-rata tertimbang model menunjukkan bahwa model tersebut berkinerja lebih baik untuk kelas 0, tetapi kinerja keseluruhannya seimbang.

4. KESIMPULAN

Model klasifikasi risiko gagal bayar kartu kredit telah dibangun menggunakan teknik *Random Forest*, dan pekerjaan ini telah berhasil menerapkannya. Menggunakan set pelatihan 70% dan set pengujian 30%, model tersebut dilatih menggunakan dataset klien kartu kredit gagal bayar yang berisi 30.000 item. Analisis kepentingan fitur menunjukkan bahwa *Pay_1*, *Age*, *Bill_Amt1*, *Limit_Bal*, *Bill_Amt2*, *Pay_Amt1*, *Bill_Amt3*, *Bill_Amt6*, *Pay_2*, *Bill_Amt4*, *Bill_Amt5*, *Pay_Amt2*, dan *Pay_Amt3* adalah elemen terpenting saat menentukan risiko gagal bayar. Faktor-faktor ini memiliki kontribusi yang signifikan dalam menentukan apakah seorang nasabah berisiko gagal bayar atau tidak. Model yang dibangun menghasilkan akurasi 82%, dengan precision 0.84 untuk kelas "tidak gagal bayar" dan 0.62 untuk kelas "gagal bayar", serta recall 0.94 untuk kelas "tidak gagal bayar" dan 0.36 untuk kelas "gagal bayar". Model ini bekerja dengan sangat baik dalam mengidentifikasi pelanggan yang tidak dalam bahaya gagal bayar (0.89 untuk kelas "tidak gagal bayar" dan 0,45 untuk kelas "gagal bayar"), tetapi masih kurang dalam mendeteksi konsumen yang berisiko gagal bayar.

REFERENCES

- [1] Y. I. Kurniawan and T. I. Barokah, 2020, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *Jurnal Ilmiah MATRIK*, vol. 22, no. 1, pp. 73–82.
- [2] M. Fauzan, 2017, "Gaya Hidup Nasabah dan Keputusan Dalam Penggunaan Kartu Kredit," *Esensi: Jurnal Bisnis dan Manajemen*, vol. 7, no. 2, pp. 181–192, doi: 10.15408/ess.v7i2.4987.
- [3] D. J. Ardha, 2020, "Analisis Kasus Pemalsuan Kartu Kredit Sebagai Bentuk Tindak Pidana Perbankan," *Jurnal Hukum Doctrinal*, vol. 5, no. 2, pp. 245–263.
- [4] H. Wihasto, 2024, "Analisis Pembiayaan Macet Pada Pembiayaan Tanpa Agunan (PTA) dengan Crediet Card (Studi Kasus Di Bank Syariah Indonesia Di Yogyakarta)," *Indonesian Journal of Innovation Science and Knowledge*, vol. 1, no. 3, pp. 208–219.
- [5] Amelda Sinaga, SS and Prof. Dr. Adler Manurung, ME.,M, Com, 2020, "Analisis Risiko 5 C Terhadap Kinerja Kartu Kredit Bca," *Jurnal Manajemen Risiko*, vol. 1, no. I, pp. 69–92, doi: 10.33541/mr.v1i1.1968.
- [6] O. Pahlevi, A. Amrin, and Y. Handrianto, 2023, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *Jurnal Infortech*, vol. 5, no. 1, pp. 71–76, doi: 10.31294/infortech.v5i1.15829.
- [7] F. A. Putri Prasetya and P. H. Prima Rosa, 2024, "Klasifikasi Kegagalan Pembayaran Kredit Nasabah Bank dengan Algoritma XGBoost," *Seminar Nasional Informatika Bela Negara (SANTIKA)*, vol. 4, no. 1, pp. 366–371.

- [8] S. F. Syah Reza and W. Cholil, 2023, "Implementasi Algoritma Random Forest Terhadap Prediksi Good Loan/Bad Loan Kredit Nasabah Bank Di Jakarta," *Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya (SENAMIKA)*, vol. 4, no. 2, pp. 535–543.
- [9] K. P. Ulandari, N. Chamidah, and A. Kurniawan, 2024, "Prediksi risiko gagal bayar kredit kepemilikan rumah dengan pendekatan metode random forest," *SAINSMAT : Jurnal Ilmiah Ilmu Pengetahuan alam*, vol. 13, no. 2, pp. 162–170.
- [10] R. Armiani and E. P. Agustini, 2022, "Analisa Fraud Pada Transaksi Kartu Kredit Menggunakan Algoritma Random Forest," *Jurnal Teknologi Informasi dan Terapan*, vol. 9, no. 2, pp. 118–126, doi: 10.25047/jtit.v9i2.297.
- [11] M. Afifudin and A. M. Rizki, 2023, "Analisis Perbandingan Penggunaan Model Machine Learning Pada Kasus Deteksi Kemampuan Calon Klien Dalam Membayar Kembali Pinjaman," *Scan : Jurnal Teknologi Informasi dan Komunikasi*, vol. 18, no. 2, pp. 32–36, doi: 10.33005/scan.v18i2.3623.
- [12] M. H. Setiono, 2022, "KOMPARASI ALGORITMA DECISION TREE, RANDOM FOREST, SVM DAN K-NN DALAM KLASIFIKASI KEPUASAN PENUMPANG MASKAPAI PENERBANGAN," *Inti Nusa Mandiri*, vol. 17, no. 1, pp. 32–39.
- [13] D. B. Saputra, V. Atina, and F. E. Nastiti, 2024, "PENERAPAN MODEL CRISP-DM PADA PREDIKSI NASABAH KREDIT MENGGUNAKAN ALGORITMA RANDOM FOREST Dwi," *Idealis: Indonesia Journal Information System Volume*, vol. 7, no. 2, pp. 240–247.
- [14] A. Primajaya and B. N. Sari, 2018, "Random Forest Algorithm for Prediction of Precipitation," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIMD)*, vol. 1, no. 1, pp. 27–31, doi: 10.24014/ijaidm.v1i1.4903.
- [15] W. Andriyani *et al.*, 2024, *PERANGKAT LUNAK DATA MINING*, Jawa Barat: Penerbit WIDINA MEDIA UTAMA.
- [16] L. Ratnawati and D. R. Sulistyningrum, 2019, "Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel," *Jurnal Sains dan Seni ITS*, vol. 8, no. 2, pp. 2337–3520, doi: 10.12962/j23373520.v8i2.48517.
- [17] A. Agung, A. Daniswara, I. Kadek, and D. Nuryana, 2023, "Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru," *Journal of Informatics and Computer Science*, vol. 05.