

PROPOSAL PENELITIAN

***ANALISIS PENDETEKSI PENIPUAN TRANSAKSI KARTU KREDIT DENGAN
ALGORITMA RANDOM FOREST***



DISUSUN OLEH:

Arya Dheffan Shevchenko

24060121140160

**DEPARTEMEN INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO**

2023

ABSTRAK

Maraknya penipuan yang sering terjadi dalam transaksi kartu kredit merupakan isu yang mendesak di era digital ini. Dengan meningkatnya penggunaan kartu kredit untuk berbagai keperluan, penipu terus mengembangkan metode mereka, mengakibatkan kerugian finansial dan kerugian kepercayaan bagi konsumen dan lembaga keuangan. Oleh karena itu, perlu adanya solusi yang dapat secara efisien dan efektif mendeteksi penipuan transaksi kartu kredit untuk mengatasi masalah ini. Penelitian ini bertujuan untuk menyediakan solusi yang dapat membantu mengurangi risiko penipuan dan melindungi keamanan transaksi kartu kredit, yang pada akhirnya akan memberikan manfaat kepada seluruh pemegang kartu kredit dan perusahaan keuangan.

Penelitian ini bertujuan untuk mengembangkan dan menganalisis sistem pendeteksi penipuan transaksi kartu kredit menggunakan algoritma *Random Forest*. Penipuan transaksi kartu kredit merupakan masalah serius dalam dunia keuangan, yang dapat menyebabkan kerugian finansial yang signifikan bagi pihak-pihak terkait. Metode *machine learning*, khususnya algoritma *Random Forest*, akan digunakan untuk memproses dan menganalisis data transaksi kartu kredit secara otomatis dengan tujuan mendeteksi transaksi yang mencurigakan. Penelitian ini akan melibatkan pengumpulan data transaksi, pemrosesan data, pelatihan model *Random Forest*, serta pengujian dan validasi model. Hasil penelitian ini diharapkan akan meningkatkan efektivitas dalam mendeteksi penipuan transaksi kartu kredit, mengurangi risiko keuangan, dan meningkatkan kepercayaan pelanggan dalam menggunakan kartu kredit mereka. Selain itu, penelitian ini juga akan menyumbang kepada perkembangan lebih lanjut dalam penerapan metode *machine learning* dalam keuangan.

Kata kunci: penipuan transaksi, *Random Forest*, Keuangan, *machine learning*

ABSTRACT

The prevalence of fraud occurring in credit card transactions is an urgent issue in this digital era. With the increasing use of credit cards for various purposes, fraudsters continue to develop their methods, resulting in financial losses and a loss of trust for consumers and financial institutions. Therefore, there is a need for a solution that can efficiently and effectively detect credit card transaction fraud to address this problem. This research aims to provide a solution that can help reduce the risk of fraud and protect the security of credit card transactions, ultimately benefiting all credit cardholders and financial institutions.

This research aims to develop and analyze a credit card transaction fraud detection system using the Random Forest algorithm. Credit card transaction fraud is a serious issue in the financial world, leading to significant financial losses for all parties involved. Machine learning methods, particularly the Random Forest algorithm, will be used to process and analyze credit card transaction data automatically with the goal of detecting suspicious transactions. This research will involve data collection, data processing, training the Random Forest model, and testing and validating the model. The results of this research are expected to enhance the effectiveness of credit card transaction fraud detection, reduce financial risks, and increase customer trust in using their credit cards. Furthermore, this research will contribute to further developments in the application of machine learning methods in the financial sector.

Keywords: *transaction fraud, Random Forest, financial, machine learning*

DAFTAR ISI

ABSTRAK	1
ABSTRACT	2
DAFTAR ISI	3
DAFTAR TABEL	4
BAB 1 PENDAHULUAN	5
1.1 Latar Belakang	5
1.2 Rumusan Masalah	6
1.3 Tujuan dan Manfaat	6
1.4 Ruang Lingkup	7
BAB II LANDASAN TEORI	9
2.1 <i>State of the Art</i>	9
2.2 Dasar Teori	12
2.2.1 <i>Credit Card Fraud Detection</i>	12
2.2.2 <i>Random Forest</i>	12
BAB III METODOLOGI PENELITIAN	13
3.1 Pengumpulan Data	13
3.2 Pengolahan Awal Data	14
3.3 Metode yang Diusulkan	15
3.4 Pengujian Metode	16
3.5 Evaluasi dan Validasi Hasil	17
BAB IV	19
5.1 Rencana Jadwal Penelitian	19
DAFTAR PUSTAKA	21
LAMPIRAN	22
5.1 <i>Problem Tree</i>	22

DAFTAR TABEL

Tabel 2.1 Penelitian Terkini Terkait Analisis Deteksi Penipuan Transaksi Kartu Kredit	9
Tabel 4.1 Rencana Jadwal Penelitian.....	19

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Menurut FBI, penipuan adalah "tindakan menipu orang lain untuk memperoleh uang, harta, atau layanan dengan menggunakan tipu daya atau kebohongan.". Penipuan dalam transaksi kartu kredit merupakan tantangan serius dalam dunia keuangan di era digital saat ini. Penipu terus mengembangkan metode mereka, yang dapat mengakibatkan kerugian finansial yang signifikan bagi konsumen, perusahaan, dan lembaga keuangan.

Menurut laporan Nilson tahun 2020 mengenai penipuan kartu kredit di seluruh dunia, kartu kredit menjadi pilihan utama karena kemudahan akses dan interaksi langsung dengan sistem perbankan. Namun, peningkatan penggunaan kartu kredit juga berarti peningkatan kasus kejahatan terkait. Dalam tahun 2019, dari total transaksi kartu kredit senilai 42.275 triliun dolar (mengalami kenaikan sebesar 4,2% dibanding tahun 2018), tercatat kerugian mencapai 28,65 milyar dolar (naik sebesar 2,9% dibanding tahun 2019).

Saat ini, machine learning merupakan komponen kunci dalam data mining dan analisis big data. Dalam konteks ini, pendekatan Random Forest (RF), yang merupakan bagian dari algoritma supervised learning pada machine learning, digunakan dalam penelitian ini untuk memberikan solusi optimal dalam mengatasi masalah deteksi penipuan. Algoritma Machine Learning digunakan untuk menganalisis semua transaksi yang terjadi dan mengidentifikasi tindakan yang mencurigakan. Laporan hasil analisis ini kemudian diperiksa oleh profesional yang akan menghubungi pemegang kartu kredit untuk memastikan apakah transaksi tersebut sah atau merupakan tindakan penipuan. Para penyelidik memberikan masukan dan umpan balik ke dalam sistem otomatis, yang kemudian digunakan untuk melatih dan memperbarui algoritma secara berkala, dengan tujuan untuk terus meningkatkan kinerja dalam mendeteksi tindakan penipuan dari waktu ke waktu.

Dalam konteks kompleksitas dan volume transaksi kartu kredit yang terus meningkat, pendekatan dengan menggunakan algoritma Random Forest jika algoritma ini dapat menawarkan solusi yang skalabel dan efektif dalam mendeteksi penipuan. Penelitian ini bertujuan untuk menjawab apakah algoritma ini dapat memproses data

transaksi dengan kecepatan dan ketepatan yang dibutuhkan, mengidentifikasi pola-pola yang mencurigakan, dan memberikan laporan yang memungkinkan profesional keamanan untuk melakukan verifikasi lebih lanjut. Hal ini tidak hanya memberikan manfaat dalam mengurangi kerugian finansial, tetapi juga meningkatkan kepercayaan pelanggan terhadap layanan kartu kredit dan perusahaan keuangan.

Selain itu, keberlanjutan peningkatan kinerja deteksi penipuan ini merupakan hal yang penting dalam menghadapi perkembangan metode penipuan yang semakin canggih. Dengan berfokus pada pengembangan dan pembaruan terus-menerus pada algoritma Machine Learning, perusahaan keuangan dapat memitigasi risiko penipuan yang dapat merugikan baik pelanggan maupun mereka sendiri. Dengan demikian, menggabungkan teknologi Machine Learning, khususnya dengan pendekatan Random Forest, menjadi strategi kunci dalam menjaga keamanan transaksi kartu kredit di era digital saat ini.

Oleh karena itu, deteksi penipuan transaksi kartu kredit telah menjadi hal yang mendesak dan penting.

1.2 Rumusan Masalah

1. Sejauh mana algoritma Random Forest efektif dalam mendeteksi penipuan dalam transaksi kartu kredit?
2. Bagaimana performa algoritma Random Forest dalam memproses dan menganalisis data transaksi kartu kredit?

1.3 Tujuan dan Manfaat

Tujuan Penelitian:

1. Menganalisis efektivitas algoritma Random Forest dalam mendeteksi penipuan dalam transaksi kartu kredit.
2. Menilai performa algoritma Random Forest dalam memproses dan menganalisis data transaksi kartu kredit.
3. Mengidentifikasi faktor-faktor yang memengaruhi tingkat keberhasilan deteksi penipuan menggunakan algoritma Random Forest.
4. Meningkatkan pemahaman tentang cara hasil analisis data transaksi dapat digunakan untuk memperbaiki sistem deteksi penipuan kartu kredit.

Manfaat Penelitian:

1. Meningkatkan keamanan transaksi kartu kredit: Penelitian ini dapat membantu meningkatkan efektivitas dalam mendeteksi penipuan transaksi kartu kredit, mengurangi risiko keuangan bagi pemegang kartu kredit, dan melindungi lembaga keuangan dari kerugian yang disebabkan oleh penipuan.
2. Peningkatan kepercayaan pelanggan: Dengan sistem deteksi penipuan yang lebih efisien, pelanggan akan lebih percaya untuk menggunakan kartu kredit mereka, menambah kepercayaan terhadap layanan perbankan.
3. Kontribusi pada perkembangan teknologi: Penelitian ini akan menyumbang kepada perkembangan lebih lanjut dalam penerapan metode machine learning, khususnya algoritma Random Forest, dalam keuangan dan sektor lainnya.
4. Penghematan sumber daya: Dengan pendekatan yang lebih baik dalam mendeteksi penipuan, perusahaan keuangan dapat menghemat sumber daya yang biasanya digunakan untuk menangani kasus penipuan.
5. Informasi lebih baik: Analisis data transaksi yang mendalam dapat memberikan wawasan yang lebih baik tentang tren penipuan yang berkembang, memungkinkan lembaga keuangan untuk mengambil tindakan pencegahan yang lebih baik di masa depan.

1.4 Ruang Lingkup

Agar permasalahan dan pembahasan lebih terarah, adapun parameter yang digunakan dalam membatasi ruang lingkup pada penelitian ini yaitu:

- a. Objek penelitian dalam studi ini adalah transaksi kartu kredit yang terjadi dalam konteks sistem perbankan atau keuangan.

- b. Penelitian ini menggunakan metode machine learning, dengan penekanan pada algoritma Random Forest. Algoritma Random Forest adalah metode supervised learning yang digunakan untuk menganalisis data transaksi kartu kredit dan mendeteksi tindakan yang mencurigakan.
- c. Penelitian ini berfokus pada analisis deteksi penipuan transaksi kartu kredit dengan menggunakan algoritma Random Forest. Fokus utama adalah untuk mengevaluasi efektivitas algoritma dalam mendeteksi penipuan, memahami faktor-faktor yang memengaruhi kinerja sistem deteksi penipuan, dan memberikan pemahaman yang dapat digunakan untuk meningkatkan sistem deteksi penipuan kartu kredit.
- d. Data yang digunakan dalam penelitian ini adalah data transaksi kartu kredit. Data ini mencakup informasi tentang setiap transaksi yang melibatkan kartu kredit, termasuk informasi tentang jumlah transaksi, tanggal dan waktu, jenis transaksi, serta data pelanggan yang terlibat. Data ini digunakan untuk melatih dan menguji algoritma Random Forest dalam mengidentifikasi tindakan penipuan.

BAB II

LANDASAN TEORI

2.1 *State of the Art*

Dalam melakukan kegiatan penelitian ini, awalnya peneliti melakukan riset terhadap beberapa penelitian terkini yang membahas mengenai analisis *User Interface* dan efektivitas serta efisiensi sebuah aplikasi atau *website* dengan berbagai metode termasuk metode yang dipakai untuk penelitian ini yaitu metode GOMS. Perbandingan penelitian atau riset terkini yang pernah dilakukan tercatat melalui Tabel 2.1.

Tabel 2.1 Penelitian Terkini Terkait Analisis *User Interface* dan *E-learning*

Peneliti	Masalah	Metode	Hasil
Farah & Maya(2023)	Analisis Random Forest Menggunakan Principal Component Analysis Pada Data Berdimensi Tinggi	Metode Random Forest	Tidak ada batasan dalam membangun $n_estimator$ pada proses klasifikasi random forest. Artinya semakin banyak $n_estimator$ yang dibangun, belum tentu menghasilkan akurasi yang baik pada proses klasifikasi dan begitu juga sebaliknya. Dan hal ini juga bergantung pada data yang digunakan. Untuk dataset airline satisfaction dan water quality akurasi meningkat pada $n_estimator=7$.
Fajar & Rahmi(2022)	Implementasi Metode Random Forest Untuk Penjurusan Siswa Di Madrasah Aliyah Negeri Sintang	Metode Random Forest	Pada setiap tree akan menghasilkan satu klasifikasi, sehingga sistem mendapatkan 5 hasil yang telah di klasifikasikan. Random forest pada sistem melakukan perhitungan hasil vote, dimana hasil nilai vote tertinggi yang menjadi hasil klasifikasi akhir. Hasil pengujian yang diperoleh menggunakan confusion matrix menghasilkan nilai accuracy sebesar 94,38%, precision sebesar 95,23%

			dan recall sebesar 94,25%.
Vinit & Rashmi(2022)	Credit Card Fraud Detection Using Support Vector Machine	Metode Support Vector Machine	Hasil menunjukkan bahwa metode SVM menghasilkan performa yang lebih bagus dilihat dari performance metrics, seperti sensitivitas, akurasi, dan specificity.
Wildan(2021)	Implementasi Metode Random Forest Pada Klasifikasi Data Ulasan Konsumen Perusahaan PT KAI	Metode random forest	Tingkat akurasi dari algoritma Random Forest dalam melakukan klasifikasi yang dihasilkan dari model yang telah dibuat adalah 88.4% yang artinya model yang telah dibuat sudah sangat baik
Darshan Kaur (2020)	Machine Learning Approach for Credit Card Fraud Detection (KNN & Naïve bayes)	Metode KNN & Naïve Bayes	Model hibrida dirancang dalam penelitian ini, yang merupakan campuran dari KNN dan klasifier naïve bayes. Model yang diusulkan diaplikasikan dalam bahasa Python dan hasilnya dibandingkan dalam hal parameter tertentu. Model yang direkomendasikan menunjukkan kinerja tinggi dibandingkan model klasifikasi berbasis pemungutan suara menurut analisis.
Mega Luna (2020)	Penerapan Analisis Random Forest Pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask	Metode <i>Random Forest</i>	Dari analisis Random Forest dengan 29 pohon, diketahui bahwa model kamera merupakan variabel dengan kepentingan terbesar yang mampu mempengaruhi prediksi harga kamera bekas dan pengaruh terkecil diberikan oleh variabel karet kamera. Dengan demikian model yang diperoleh untuk memprediksi harga kamera bekas memiliki nilai akurasi sebesar 76.7%.

Berdasarkan state-of-the-art yang telah disajikan, terdapat beberapa penelitian yang menyoroti keefektifan metode Random Forest dalam konteks analisis data. Misalnya, penelitian oleh Fajar & Rahmi (2022) menunjukkan bahwa Random Forest dapat memberikan hasil klasifikasi yang baik dengan tingkat akurasi yang tinggi dalam penjurusan siswa. Begitu pula, penelitian oleh Wildan (2021) menunjukkan tingkat akurasi yang signifikan, mencapai 88,4%, dalam klasifikasi data ulasan konsumen menggunakan Random Forest. Meskipun hasil tersebut tidak secara langsung terkait dengan deteksi penipuan kartu kredit, mereka memberikan indikasi bahwa Random Forest dapat memberikan kinerja yang baik dalam tugas klasifikasi.

Selanjutnya, keberhasilan metode Random Forest juga tercermin dalam penelitian Mega Luna (2020) terkait prediksi harga kamera bekas. Meskipun akurasi sebesar 76,7% mungkin lebih rendah dibandingkan dengan penelitian lainnya, hal ini tetap menunjukkan kemampuan model dalam mengatasi kompleksitas data. Dengan demikian, berdasarkan informasi ini, dapat diasumsikan bahwa Random Forest memiliki potensi untuk menjadi metode yang efektif dalam menganalisis deteksi penipuan kartu kredit. Kelebihan Random Forest dalam mengatasi kompleksitas dan variasi data, seperti yang terlihat dalam penelitian-penelitian tersebut, membuatnya menjadi pilihan yang menjanjikan untuk tugas-tugas klasifikasi yang melibatkan deteksi anomali atau penipuan.

2.2 Dasar Teori

2.2.1 *Credit Card Fraud Detection*

Deteksi penipuan kartu kredit merupakan aplikasi kritis dari teknologi kecerdasan buatan (AI) dan machine learning (ML) dalam industri keuangan. Tujuan utamanya adalah mengidentifikasi transaksi yang mencurigakan atau tidak sah, memungkinkan lembaga keuangan dan pemegang kartu untuk mengambil langkah-langkah pencegahan secepat mungkin. Metode machine learning, seperti yang terlihat dalam berbagai penelitian, memainkan peran kunci dalam mencapai tujuan ini. Random Forest, salah satu metode yang umum digunakan, melibatkan pembangunan beberapa pohon keputusan dan penggabungan hasilnya untuk meningkatkan akurasi deteksi.

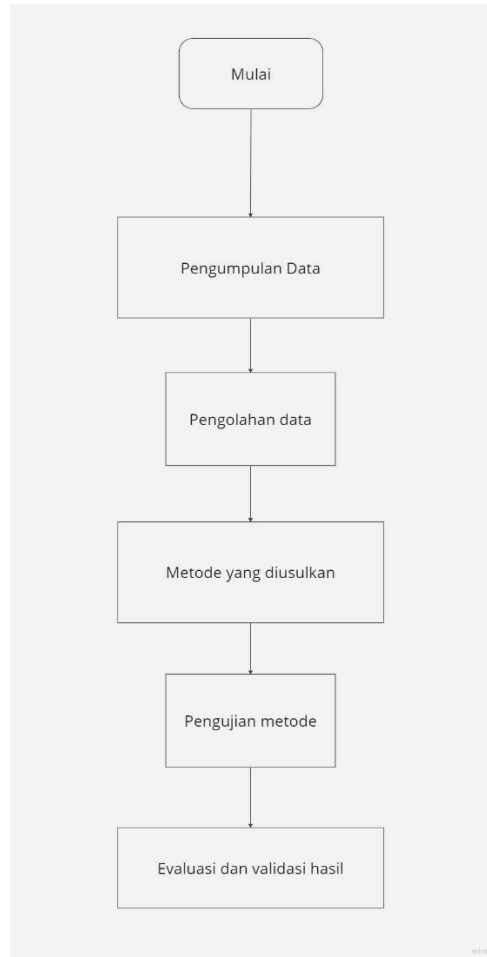
2.2.2 *Random Forest*

Random Forest adalah sebuah metode dalam machine learning yang beroperasi dengan prinsip ensemble learning, di mana sejumlah besar pohon keputusan digunakan untuk meningkatkan keakuratan dan kestabilan prediksi. Setiap pohon keputusan dibangun menggunakan subset acak dari data, yang dikenal sebagai bootstrapped samples, yang diperoleh dengan pengambilan sampel dengan penggantian dari data yang ada. Pentingnya penggunaan bootstrapped sampling adalah untuk menciptakan variasi antar pohon, menghindari overfitting, dan membuat model lebih robust. Selain itu, Random Forest juga melibatkan feature randomization, yaitu ketika hanya sebagian kecil dari fitur yang dipertimbangkan saat membangun setiap pohon. Hasil prediksi dari setiap pohon diambil, dan melalui mekanisme voting atau averaging, prediksi akhir dibuat. Dengan pendekatan ini, Random Forest mampu mengatasi kompleksitas data, meningkatkan kinerja model, dan seringkali memberikan hasil yang lebih baik daripada pohon keputusan tunggal. Kelebihan ini membuatnya menjadi pilihan populer dalam berbagai tugas, termasuk klasifikasi dan regresi.

BAB III

METODOLOGI PENELITIAN

Metodologi penelitian di sini memiliki beberapa tahap atau langkah seperti yang tercantum di Gambar 3.1.



Gambar 3.1 Diagram Tahapan Metodologi Penelitian

Mengacu Gambar 3.1 memperlihatkan mengenai tahap atau langkah pada penelitian ini, dimulai dari pengumpulan data, lalu pengolahan data yang dilanjutkan dengan metode yang diusulkan, selanjutnya pengujian metode, kemudian evaluasi dan validasi hasil yang setelah itu menandakan proses metodologi penelitian selesai.

3.1 Pengumpulan Data

Data penipuan kartu kredit dapat diperoleh di website bernama kaggle. Kaggle adalah platform online yang menyediakan berbagai kompetisi data, sumber daya pembelajaran, dan alat untuk para ilmuwan data dan praktisi machine learning. Didirikan pada tahun 2010, Kaggle bertujuan

untuk menyatukan komunitas data science dan machine learning dengan memberikan tantangan (competitions) di mana para pesertanya dapat bersaing untuk memecahkan masalah analisis data tertentu.

3.2 Pengolahan Awal Data

Pre-processing data sangat penting dalam analisis data dan pembuatan model machine learning. Data Preprocessing merupakan salah satu tahapan dalam melakukan mining data. Sebelum menuju ke tahap pemrosesan, Data mentah akan diolah terlebih dahulu. Data Preprocessing atau praproses data biasanya dilakukan melalui cara eliminasi data yang tidak sesuai. Selain itu dalam proses ini data akan diubah dalam bentuk yang akan lebih dipahami oleh sistem.

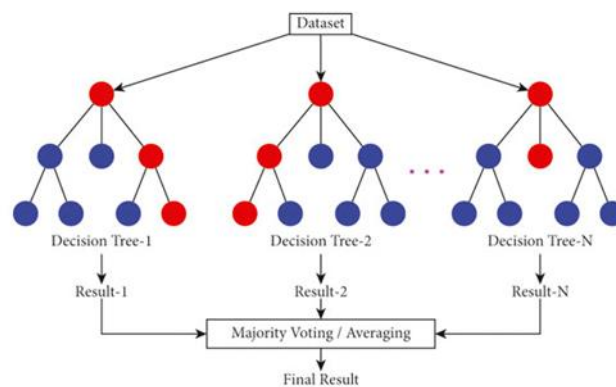


Gambar 3.2. Tahap Preprocessing

Pada ilustrasi di atas dapat dijelaskan bahwa proses preprocessing data dimulai dengan tahap cleaning, di mana dataset diperiksa untuk mengidentifikasi dan menangani nilai yang hilang, outliers, dan duplikat. Setelah itu, langkah integrasi melibatkan penggabungan sumber data jika data berasal dari beberapa sumber. Transformation data melibatkan normalisasi, encoding variabel kategorikal, dan transformasi fitur untuk menyamakan rentang nilai dan format data.

Reduksi baris data adalah proses penyederhanaan dataset dengan tujuan membuatnya lebih mudah diinterpretasi atau diolah. Ada beberapa metode yang dapat digunakan untuk mencapai ini. Salah satu pendekatan umum adalah pengambilan sampel, yang dapat dilakukan secara acak atau berdasarkan kelompok tertentu. Pengambilan sampel dapat membantu mengurangi jumlah baris tanpa mengorbankan representativitas dataset. Metode lain melibatkan penyaringan data dengan menghilangkan baris yang tidak memenuhi kriteria tertentu, seperti kondisi filter atau nilai outlier. Agregasi adalah teknik lain yang melibatkan penggabungan beberapa baris ke dalam satu baris berdasarkan kelompok tertentu atau menggunakan fungsi agregasi.

3.3 Metode yang Diusulkan



Gambar 3.3 Diagram Skema Metode yang Diusulkan

Gambar 3.3 Random forest adalah metode pembelajaran mesin ensemble yang menggunakan banyak decision tree untuk membuat prediksi. Tahap-tahap random forest adalah sebagai berikut:

- **Pemilihan data**
Data yang akan digunakan untuk pelatihan random forest harus dipersiapkan terlebih dahulu. Data tersebut harus dibersihkan dari noise dan outlier, serta dinormalisasi agar nilainya berada dalam rentang yang sama.
- **Pembuatan decision tree**
Decision tree adalah struktur pohon yang digunakan untuk membuat

keputusan berdasarkan serangkaian aturan. Decision tree dibuat dengan menggunakan algoritma greedy yang memilih fitur yang paling informatif pada setiap langkah.

- Pembuatan kumpulan decision tree

Pada tahap ini, sejumlah decision tree dibuat dengan menggunakan teknik bagging. Bagging adalah teknik yang digunakan untuk membuat sejumlah model dari data yang sama dengan menggunakan bootstrapping. Bootstrapping adalah teknik yang digunakan untuk membuat data baru dengan cara mengambil sampel data asli secara acak dengan penggantian.

- Pembuatan prediksi

Prediksi dibuat dengan menggunakan teknik voting atau averaging. Voting adalah teknik yang digunakan untuk menentukan hasil prediksi berdasarkan suara terbanyak dari sejumlah decision tree. Averaging adalah teknik yang digunakan untuk menentukan hasil prediksi berdasarkan rata-rata dari sejumlah decision tree.

3.4 Pengujian Metode

Menguji metode Random Forest di Jupyter Notebook melibatkan beberapa langkah. Di bawah ini adalah panduan umum tentang cara melakukannya:

1. Instalasi Paket:

Pastikan kita telah menginstal paket-paket yang diperlukan seperti scikit-learn untuk mengakses implementasi Random Forest dan pandas atau numpy untuk manipulasi data. Jika belum diinstal, kita dapat melakukannya dengan perintah berikut di Jupyter Notebook:

2. Impor Paket:

Impor paket-paket yang dibutuhkan di awal notebook kita. Di antara paket-paket tersebut biasanya termasuk ``numpy``, ``pandas``, dan ``sklearn`` untuk Random Forest.

3. Baca dan Persiapkan Data:

Baca data Kita (misalnya, dari file CSV) dan lakukan pra-pemrosesan data

seperti mengisi nilai yang hilang, mengonversi variabel kategorikal menjadi numerik, dan membagi data menjadi set pelatihan dan pengujian.

4. Inisialisasi dan Pelatihan Model:

Inisialisasikan model Random Forest dan latih model menggunakan set pelatihan.

5. Evaluasi Model:

Evaluasi model menggunakan set pengujian dan metrik evaluasi yang relevan seperti akurasi, presisi, dan recall.

6. Optimasi dan Penyetelan Parameter (Opsional):

Jika perlu, kita dapat melakukan optimasi dan penyetelan parameter untuk meningkatkan kinerja model. Ini dapat melibatkan penyetelan jumlah pohon (`n_estimators`), kedalaman pohon (`max_depth`), dan parameter lainnya.

7. Visualisasi Hasil (Opsional):

Visualisasikan hasil atau kinerja model jika diperlukan, seperti fitur yang paling penting atau kurva ROC.

Pastikan untuk menggantikan ``nama_file.csv``, ``features``, dan ``target`` dengan nilai yang sesuai dengan dataset. Langkah-langkah ini memberikan gambaran umum tentang bagaimana menguji metode Random Forest di Jupyter Notebook, dan dapat menyesuaikannya dengan kebutuhan spesifik proyek kita.

3.5 Evaluasi dan Validasi Hasil

Untuk evaluasi model Random Forest, kita dapat menggunakan berbagai metrik evaluasi tergantung pada tipe masalah yang sedang kita hadapi. Beberapa metrik evaluasi umum untuk model klasifikasi (classification) melibatkan perbandingan antara prediksi model dan nilai sebenarnya dari data pengujian. Berikut adalah beberapa metrik evaluasi umum yang dapat digunakan:

1. Akurasi (Accuracy):

Akurasi mengukur sejauh mana model dapat memprediksi dengan benar seluruh

kelas. Ini dinyatakan sebagai persentase prediksi yang benar dari total prediksi.

```
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_true, y_pred)
```

2. Presisi (Precision):

Presisi mengukur sejauh mana prediksi positif model adalah benar. Ini dinyatakan sebagai rasio positif benar terhadap total positif yang diprediksi.

```
from sklearn.metrics import precision_score
precision = precision_score(y_true, y_pred)
```

3. Recall (Sensitivitas atau True Positive Rate):

Recall mengukur sejauh mana model dapat mendeteksi seluruh kelas positif. Ini dinyatakan sebagai rasio positif benar terhadap total positif yang sebenarnya.

```
from sklearn.metrics import recall_score
recall = recall_score(y_true, y_pred)
```

4. F1-Score:

F1-score adalah ukuran gabungan dari presisi dan recall dan berguna ketika ada ketidakseimbangan antara kelas positif dan negatif.

```
from sklearn.metrics import f1_score
f1 = f1_score(y_true, y_pred)
```

5. Confusion Matrix:

Matriks kebingungan (confusion matrix) memberikan pandangan rinci tentang performa model, termasuk True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_true, y_pred)
```

BAB IV

RENCANA JADWAL PENELITIAN

5.1 Rencana Jadwal Penelitian

Jadwal ini dirancang untuk keberjalanan penelitian agar lebih terstruktur. Berdasarkan jadwal yang ada, peneliti sangat berharap besar bahwa keberjalanannya akan sesuai dengan jadwal yang telah dirancang atau direncanakan. Jadwal penelitian tertera melalui Tabel 4.1.

Tabel 4.1 Rencana Jadwal Penelitian

No	Jenis Kegiatan	Nov 2023				Dec 2023				Januari 2023				Febuari 2023				Maret 2023			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan Data																				
2	Pengolahan Awal Data																				
3	Pengusulan Metode																				
4	Pengujian Metode																				
5	Evaluasi dan Validasi Hasil																				

Tabel 4.1 menjelaskan mengenai rangkaian jadwal penelitian yang dimulai dari pengumpulan data dilakukan mulai minggu pertama November. Lalu dilanjutkan dengan pengolahan awal data yang dilakukan selama 7 minggu mulai minggu ke-2 November hingga minggu ke-4 Desember. Setelah pengolahan awal data, berlanjut ke pengusulan metode dilakukan sejak minggu ke-3 Desember sampai minggu ke-2 Januari. Selanjutnya adalah pengujian metode yang membutuhkan waktu cukup

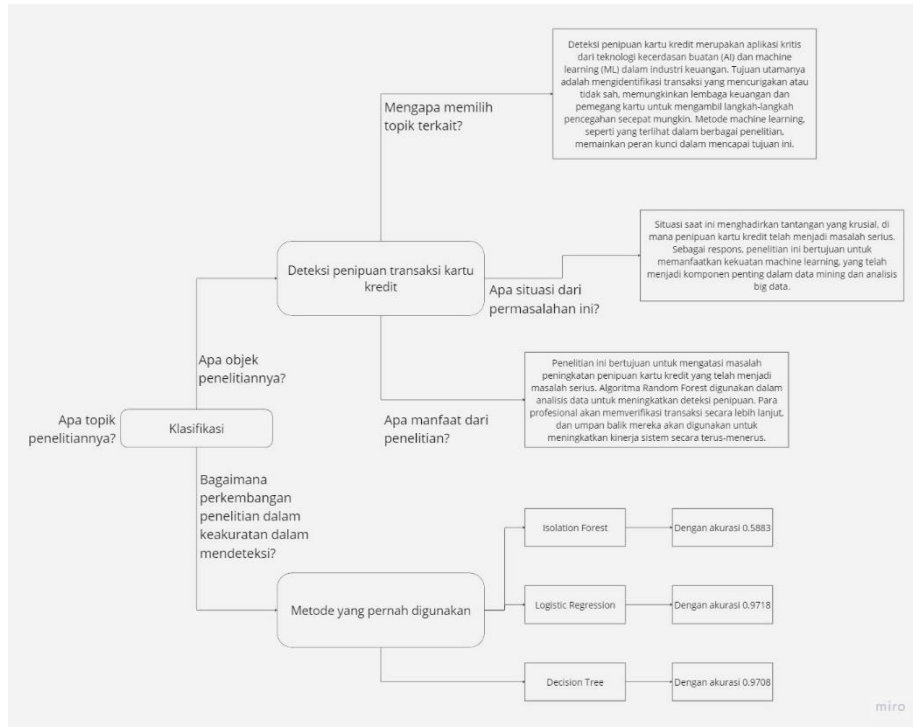
lama yaitu dari minggu ke-4 Desember hingga minggu ke-1 Maret. Terakhir, evaluasi dan validasi hasil yang dilakukan sejak minggu ke-2 sampai minggu ke-4 Maret.

DAFTAR PUSTAKA

- Arief Kurniawan & Yulianingsih (2021). VORI: Pendugaan Fraud Detection pada kartu kredit dengan Machine Learning. <https://doi.org/10.33322/kilat.v10i2.1482>
- Carcillo, F., Le Borgne, Y., Caelen, O., & Bontempi, G. (2018). Streaming Active Learning Strategies for Real-Life Credit Card Fraud Detection: Assessment and Visualization. <https://doi.org/10.48550/arXiv.1804.07481>
- Darshan Kaur & Shubpreet Kaur(2020). Machine Learning Approach for Credit Card Fraud Detection (KNN & Naïve bayes)
- Farah Diba & Maya Silvi(2023). Analisis Random Forest Menggunakan Principal Component Analysis Pada Data Berdimensi Tinggi. <https://doi.org/10.33022/ijcs.v12i4.3329>
- Mega Luna(2020). Penerapan Analisis Random Forest Pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask.
- Putu Tirta(2022). Analisis Sistem Pendeteksi Penipuan Transaksi Kartu Kredit dengan Algoritma Machine Learning. <https://doi.org/10.37012/jtik.v8i2.1306>
- Ramayanti, P. & Sutabri, S. (2023). Analisis Pendeteksi Penipuan Transaksi Kartu Kredit Dengan Algoritma Naïve Bayes dan Support Vector Machine. Jurnal Informatika Teknologi dan Sains, 5(1). <https://doi.org/10.2686-3359/online/vol5/no1/februari2023>
- Trivusi(2022). Algoritma Random Forest: Pengertian dan Kegunaannya. <https://www.trivusi.web.id/2022/08/algoritma-random-forest.html>
- Vinit Kumar & Rashmi Pathak(2022). Credit Card Fraud Detection Using Support Vector Machine. DOI: 10.1007/978-981-16-6407-6_3
- Wildan(2021). Implementasi Metode Random Forest Pada Klasifikasi Data Ulasan Konsumen Perusahaan PT KAI

LAMPIRAN

5.1 Problem Tree



Gambar 5.1 Problem Tree

Gambar 5.1 menjelaskan mengenai *problem tree* dari penelitian ini yang mana berisi topik penelitian, objek penelitian, mengapa memilih topik atau bahasan yang diangkat dalam penelitian ini, dampak dari adanya masalah tersebut, dan manfaat dari penelitian yang dilakukan. Terdapat pula metode yang pernah digunakan beserta akurasi sehingga pada akhirnya memilih metode yang paling efektif untuk pengujian masalah dalam penelitian ini.

