

Nama : Revani Aprilia
NIM : 23283030
Semester : 4
Mata Kuliah : Kecerdasan Buatan (AI)

Jawaban

Judul : Deteksi Penipuan Kartu Kredit

Deskripsi Masalah

Deteksi penipuan kartu kredit adalah suatu proses untuk mengidentifikasi suatu transaksi yang mencurigakan yang mungkin merupakan suatu kasus kategori penipuan. Dalam kasus ini, kita juga bisa menggunakan suatu pendekatan klasifikasi untuk dapat menentukan apakah suatu transaksi tersebut adalah Fraud (penipuan) atau Non-Fraud (normal). Masalah ini sangat penting dalam dunia nyata karena kasus penipuan kartu kredit dapat menyebabkan suatu kerugian finansial besar bagi pengguna maupun Perusahaan, oleh karena itu, deteksi dini terhadap transaksi mencurigakan juga sangatlah penting.

Jenis Masalah

Klasifikasi : Bukan Regresi
Tujuan : Mengklasifikasikan transaksi ke dalam dua kelas :
➤ 0 : Transaksi Normal
➤ 1 : Transaksi Penipuan

Dataset

Dataset : Credit Card Fraud Detection
Sumber : Kaggle_Credit_Card_Fraud_Detection
Ukuran : 284.807 data transaksi
Fitur :
➤ Fitur hasil PCA : V1, V2, ..., V28 (untuk alasan privasi)
➤ Amount : Nominal transaksi
➤ Time : Waktu transaksi
➤ Class : Target label (0/1)

Metode / Algoritma

Menggunakan metode algoritma : **Naive Bayes Classifier**

- Tipe : Gaussian Naive Bayes (karena fiturnya numerik)
- Asumsi : Antar fitur bersifat independent
- Cocok untuk klasifikasi cepat dan bekerja cukup baik untuk Dataset besar

Tahapan Penyelesaian Masalah

1. Persiapan & Import Library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score
```

- Digunakan library untuk analisis data, visualisasi, preprocessing, dan modeling.

2. Memuat Dataset

```
data = pd.read_csv('creditcard.csv', sep=',')
data.head()
```

- Dataset transaksi kartu kredit dimuat untuk dianalisis.
- Terdapat kolom Time, Amount, dan Class (label 1 = fraud, 0 = normal). Dataset transaksi kartu kredit dimuat untuk dianalisis.

3. Preprocessing

```
df['Amount'] = StandardScaler().fit_transform(df['Amount'].values.reshape(-1,1))
df = df.drop(['Time'], axis=1)
```

- Melihat distribusi kelas (class)
- Visualisasi kolerasi antar fitur
- Mengecek apakah dataset tidak seimbang (imbalanced)

- Kolom Amount dinormalisasi.
- Kolom Time dihapus karena kurang relevan.

4. Eksplorasi Data Awal

```
data.info()  
data.isnull().values.any()
```

- Mengecek tipe data dan memastikan tidak ada missing value.

5. Pemisahan Data Fraud & Normal

```
fraud = data[data['Class']==1]  
  
normal = data[data['Class']==0]
```

- Data dikelompokkan untuk membandingkan transaksi fraud dan normal.

6. Analisis Statistik Kolom

```
fraud.Amount.describe()  
normal.Amount.describe()
```

- Melihat perbedaan distribusi jumlah uang antara transaksi fraud dan normal.

7. Visualisasi Distribusi Jumlah Uang

```
plt.hist(fraud.Amount), plt.hist(normal.Amount)
```

- Menampilkan histogram jumlah uang yang digunakan tiap kelas.

8. Pengambilan Sampel (Sampling)

```
data1 = data.sample(frac=0.1, random_state=1)
```

- Dataset diperkecil menjadi 10% untuk mempercepat proses training.

9. Menentukan Rasio Outlier

```
outlier_fraction = len(Fraud)/float(len(Valid))
```

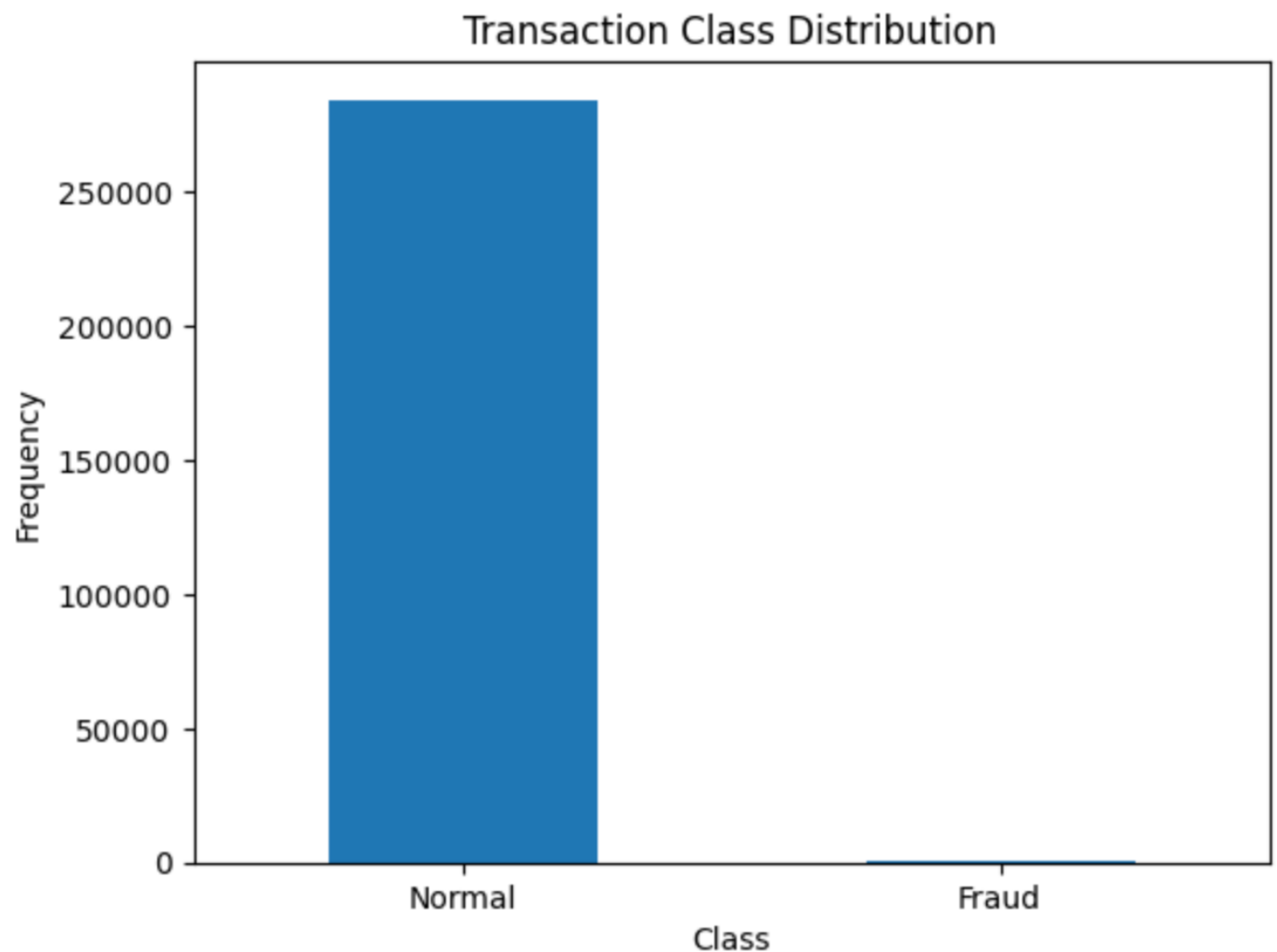
- Menghitung proporsi fraud terhadap transaksi valid.

10. Korelasi Fitur

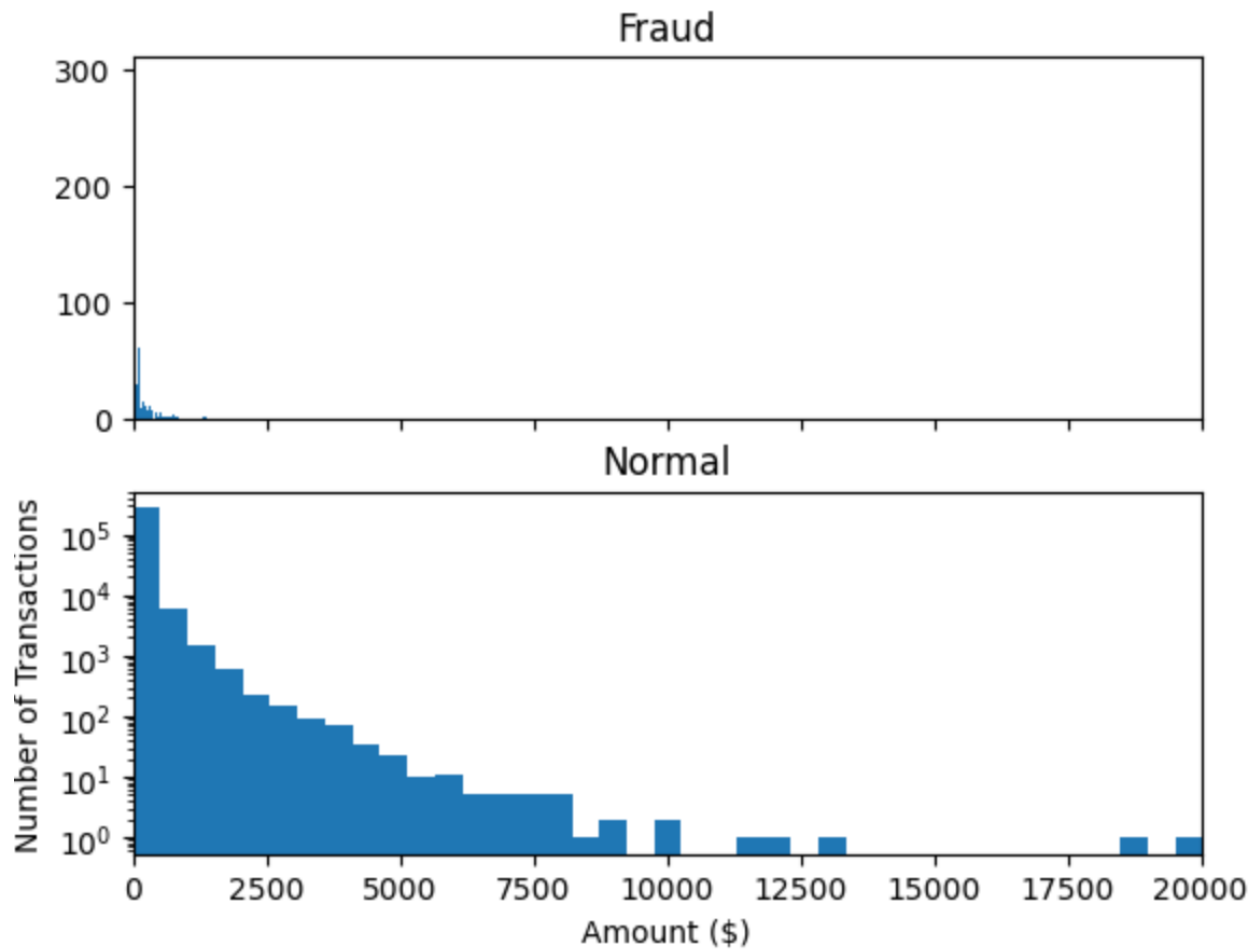
```
corr_matrix = df.corr()  
cor_target_sorted = corr_matrix["Class"].drop("Class").abs().sort_values(ascending=False)
```

- Menampilkan fitur yang paling berkorelasi terhadap penipuan.

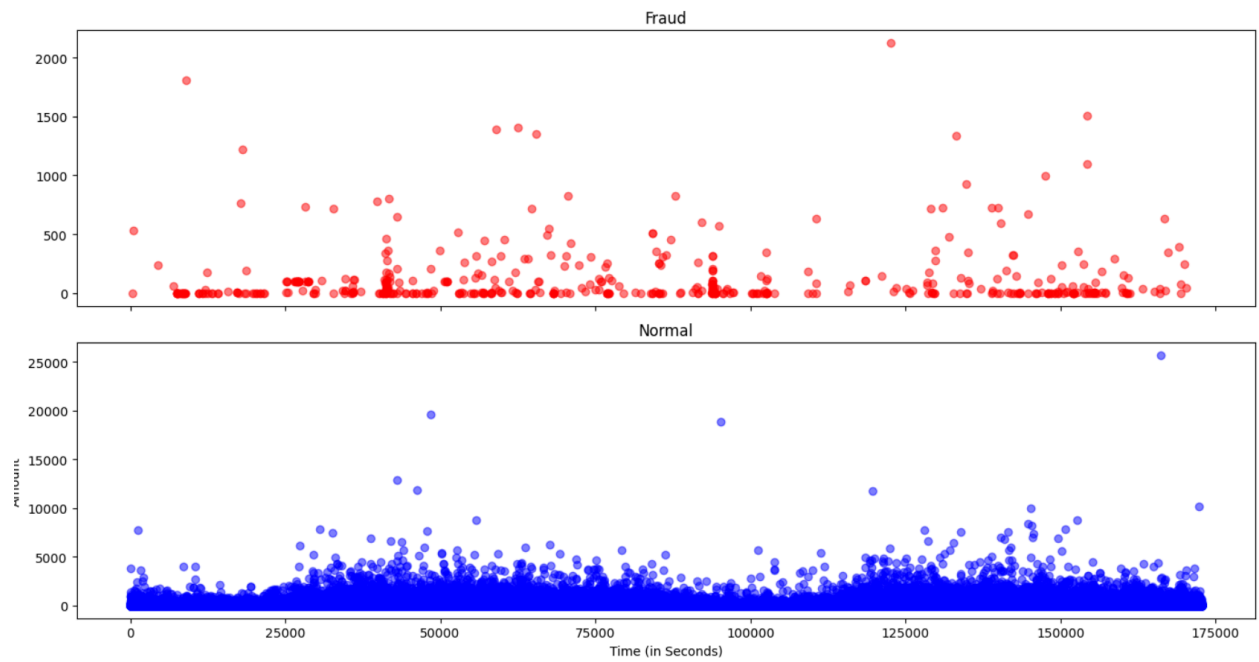
Hasil



Amount per transaction by class



Time of Transaction vs Amount by Class



Kesimpulan

1. Dataset yang digunakan adalah data transaksi kartu kredit dengan label penipuan (fraud) dan normal.
2. Preprocessing data dilakukan dengan menghapus kolom yang tidak relevan (Time) dan menormalisasi kolom Amount.
3. Eksplorasi data awal menunjukkan distribusi yang sangat tidak seimbang antara transaksi normal dan fraud.
4. Analisis statistik membuktikan bahwa transaksi fraud memiliki pola jumlah uang yang berbeda dari transaksi normal.
5. Sampling diambil untuk efisiensi proses analisis, dan rasio penipuan dihitung sebagai referensi.
6. Korelasi fitur dilakukan untuk mengetahui atribut yang paling relevan dalam mendeteksi penipuan.
7. Dataset sudah siap untuk pemodelan machine learning tahap selanjutnya, seperti klasifikasi menggunakan Random Forest.

Referensi

<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>