

LAPORAN TUGAS AWAL (UTS)

DATA MINING

DOSEN PENGAMPU : DR. ARYA ADHYAKSA WASKITA, S.SI., M.SI.



Disusun Oleh :

Nama : Aditya Eko Nugroho

Nim : 231011401122

Kelas : 04TPLP027 – REG A

TEKNIK INFORMATIKA S1
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PAMULANG

JL. RAYA PUSPITEK, BUARAN, KEC. PAMULANG, KOTA TANGERANG SELATAN
BANTEN. 15560

2025

LAPORAN TUGAS DATA MINING

Judul : Analisis Trafik Jaringan Normal Menggunakan Dataset “Benign Traffic”

1. Pendahuluan

Dalam era digital saat ini, keamanan jaringan menjadi aspek krusial dalam pengelolaan sistem informasi. Data mining, sebagai teknik untuk mengekstraksi informasi berharga dari kumpulan data besar, dapat digunakan untuk menganalisis pola trafik jaringan. Analisis ini membantu dalam memahami perilaku normal jaringan, yang selanjutnya dapat digunakan sebagai dasar untuk mendeteksi anomali atau serangan siber. Dalam tugas ini, kami menggunakan dataset "Benign Traffic" yang merepresentasikan lalu lintas jaringan normal untuk memahami karakteristik trafik yang tidak berbahaya.

2. Tujuan

- Mengeksplorasi dan memahami struktur dataset "Benign Traffic".
- Melakukan analisis statistik dan visualisasi data untuk mengidentifikasi pola trafik normal.
- Menyiapkan data untuk tahap analisis lanjutan dalam deteksi anomali jaringan.

3. Alat dan Teknologi

- Python 3
- Pandas (manipulasi data)
- Matplotlib & Seaborn (visualisasi data)
- Scikit-learn (pra-pemrosesan dan evaluasi model)

4. Langkah-langkah Analisis

a. Import Library

```
import pandas as pd
```

```
import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler
```

b. Membaca Dataset

```
df = pd.read_csv("Benign Traffic.csv")
```

c. Eksplor Data

```
print(df.head())

print(df.info())

print(df.describe())
```

d. Visualisasi Data

```
plt.figure(figsize=(10,8))

sns.heatmap(df.corr(), annot=True, cmap="coolwarm")

plt.title("Korelasi Antar Fitur")

plt.show()
```

5. Kendala Yang Dihadapi

- Beberapa library belum terinstal, seperti matplotlib dan seaborn.
- Dataset tidak memiliki label untuk klasifikasi, sehingga analisis terbatas pada eksplorasi data.
- Beberapa fitur memiliki nilai yang sangat bervariasi, memerlukan normalisasi untuk analisis lebih lanjut.

6. Kendala dan Solusi

Beberapa fitur tidak relevan

- Solusi: Fitur yang tidak berpengaruh bisa dibuang dengan fungsi `drop()` di pandas.

```
df = df.drop(columns=["Unnamed: 0", "Flow ID", "Timestamp"], errors='ignore')
```

Dataset tidak memiliki label target

► Solusi: Untuk analisis supervised, bisa gabungkan dataset ini dengan dataset lain yang memiliki label. Atau gunakan metode unsupervised seperti clustering (KMeans).

7. Kesimpulan

Melalui praktikum ini, mahasiswa mendapatkan pemahaman awal mengenai cara mengakses, mengevaluasi, dan memvisualisasikan dataset lalu lintas jaringan yang bersifat normal. Dataset "*Benign Traffic*" telah berhasil ditelaah dari sisi struktur data, statistik deskriptif, serta korelasi antar fitur. Tahapan ini menjadi landasan penting dalam proses data mining, khususnya dalam membedakan data normal dan anomali pada jaringan.