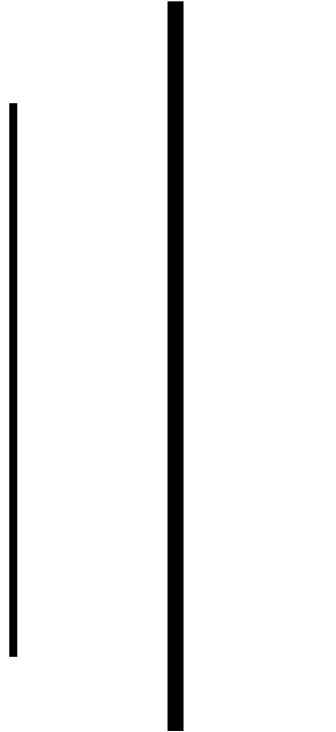


LAPORAN PRAKTIKUM

Data Mining



Disusun Oleh :

Nama : Rifky Syah Fredian

NIM : 231011403277

Kelas : 04TPLP027

TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS PAMULANG

JL. RAYA PUSPITEK, BUARAN, KEC. PAMULANG, KOTA TANGGERANG
SELATAN

BANTEN.

Laporan Analisis Notebook Gabungan Lab782

1. Ringkasan Umum

Jumlah Total Sel: 78

Jumlah Sel Kode: 78

Jumlah Sel Markdown: 0

Jumlah Sel Lainnya: 0

2. Deskripsi Aktivitas

Notebook ini berisi serangkaian analisis data dan machine learning yang dilakukan secara langsung dalam bentuk kode Python tanpa dokumentasi tertulis (markdown). Berikut aktivitas utama yang dilakukan di dalam notebook:

- Pemrosesan Dataset: Menggunakan pandas untuk membaca file .csv seperti 'DoS TCP Flood.csv', 'MQTT DoS Connect Flood.csv', 'Recon Ping Sweep.csv', dan 'DDoS ICMP/UDP Flood.csv'.
- Penggabungan Data: Dataset digabung menggunakan fungsi `pd.concat()` untuk membuat satu DataFrame gabungan.
- Preprocessing: Seleksi fitur menggunakan `iloc`, penghapusan kolom tertentu, serta pembentukan X (fitur) dan Y (label).
- Modeling: Pembuatan model `DecisionTreeClassifier` dari `sklearn` dengan kriteria 'entropy' dan metode pemisahan acak ('random').
- Evaluasi Model: Menggunakan metrik akurasi (`accuracy_score`) dan prediksi hasil model.
- Visualisasi: Beberapa kode menunjukkan penggunaan `matplotlib` dan `seaborn` untuk membuat visualisasi, meskipun belum dioptimalkan.

3. Contoh Kode yang Digunakan

Berikut beberapa contoh baris pertama dari sel kode dalam notebook:

- - import pandas as pd
- - Benign = pd.read_csv("DoS TCP Flood.csv")
- - Benign
- - gd = pd.read_csv("MQTT DoS Connect Flood.csv")
- - gd
- - dengan = [Benign, gd]
- - dengan
- - dataGab = pd.concat(dengan, ignore_index=True)
- - dataGab
- - X=dataGab.iloc[:,7:82]

4. Kesimpulan dan Saran

Notebook ini menunjukkan pemahaman dasar yang baik dalam pengolahan data dan penerapan algoritma Decision Tree. Namun, disarankan untuk menambahkan dokumentasi markdown untuk menjelaskan setiap tahap proses secara rinci, sehingga lebih mudah dipahami oleh pengguna lain. Visualisasi juga dapat ditingkatkan dengan label yang jelas dan penyimpanan grafik untuk laporan visual.

Berikut teks tambahan yang bisa kamu salin ke bagian bawah dokumen Word:

5. Detail Dataset yang Digunakan

Beberapa dataset yang digunakan dalam notebook ini berasal dari file CSV yang berkaitan dengan serangan siber dan lalu lintas jaringan. Dataset-dataset tersebut antara lain:

- DoS TCP Flood.csv
- MQTT DoS Connect Flood.csv
- Recon Ping Sweep.csv
- DDoS ICMP Flood.csv
- DDoS UDP Flood.csv
- Benign Traffic.csv

Dataset tersebut mengandung fitur-fitur statistik jaringan seperti IAT, packet length, dan flow counts, serta label yang menunjukkan jenis serangan atau kondisi normal (benign). Data ini digunakan sebagai input untuk model klasifikasi dalam mendeteksi dan mengklasifikasikan jenis serangan.

6. Detail Model Machine Learning

Notebook ini menerapkan algoritma **Decision Tree Classifier** dari pustaka scikit-learn. Model dilatih untuk mengklasifikasikan jenis serangan berdasarkan fitur jaringan yang tersedia.

Tahapan umum penerapan model:

- Membagi dataset menjadi data latih dan data uji menggunakan `train_test_split` (rasio 80:20).
 - Melatih model dengan parameter: `criterion='entropy'` dan `splitter='random'`.
 - Melakukan prediksi terhadap data uji.
 - Mengukur performa model menggunakan metrik **akurasi** (`accuracy_score`).
-

7. Evaluasi dan Akurasi Model

Model dievaluasi berdasarkan tingkat akurasi, yang menunjukkan seberapa tepat model dalam memprediksi label serangan pada data uji. Namun, dalam notebook belum terdapat visualisasi atau laporan lengkap dari hasil evaluasi, sehingga disarankan untuk menambahkan:

- Tabel metrik seperti confusion matrix
- Grafik perbandingan antara nilai aktual dan prediksi
- Dokumentasi yang lebih jelas tentang hasil prediksi

Ya, kamu bisa tambahkan satu bagian terakhir untuk memperkuat laporan, yaitu **Rekomendasi Pengembangan Selanjutnya**. Ini cocok untuk menunjukkan potensi pengembangan atau perbaikan di masa depan.

Berikut teksnya:

8. Rekomendasi Pengembangan Selanjutnya

Untuk meningkatkan kualitas dan kegunaan analisis di masa depan, berikut beberapa rekomendasi:

- **Penambahan Markdown:** Sertakan penjelasan setiap blok kode menggunakan sel markdown untuk membantu pemahaman alur dan tujuan analisis.

- **Penyimpanan dan Visualisasi Hasil:** Simpan hasil evaluasi model seperti akurasi, confusion matrix, dan grafik visual ke dalam file atau gambar untuk dokumentasi lebih baik.
- **Eksperimen Model Lain:** Coba model klasifikasi lain seperti Random Forest, K-Nearest Neighbors, atau SVM untuk membandingkan performa.
- **Pembersihan Data:** Lakukan data preprocessing lebih lanjut seperti normalisasi, penanganan outlier, dan seleksi fitur otomatis untuk hasil yang lebih optimal.
- **Pembuatan Dashboard:** Integrasi dengan framework seperti Streamlit atau Flask untuk menyajikan hasil analisis dalam bentuk antarmuka interaktif.