

UTS PEMBELAJARAN MESIN



CLASSIFICATION MODEL

NICHOLAS ALVITO DIANDRA

1103210207

TEKNIK KOMPUTER

IMPORT PANDAS

```
# Import library yang diperlukan
import pandas as pd

# Memuat dataset
file_path = '/content/data 3.csv' # Ganti dengan nama file dataset Anda
data = pd.read_csv(file_path) # Membaca file CSV ke dalam DataFrame

# Menampilkan beberapa baris awal dataset untuk memahami struktur data
data.head() # Menampilkan 5 baris pertama dari dataset
```

	android.permission.GET_ACCOUNTS	com.sonyericsson.home.permission.BROADCAST_BADGE	android.permission.READ_PROFILE	android.permission.MANAGE_ACCOUNTS	android.permission.WRI
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

5 rows × 87 columns

FUNGSI INI MENGIMPOR PUSTAKA PANDAS, MEMBACA FILE CSV DARI PATH YANG DITENTUKAN, DAN MENAMPILKAN 5 BARIS PERTAMA UNTUK MEMBERI GAMBARAN TENTANG STRUKTUR DATA.

```
# Informasi dasar dataset
data_info = {
    "Shape": data.shape, # Menampilkan dimensi dataset (jumlah baris dan kolom)
    "Missing Values": data.isnull().sum().sum(), # Menghitung total nilai yang hilang
    "Duplicate Rows": data.duplicated().sum(), # Menghitung jumlah baris duplikat
    "Target Value Counts": data['Result'].value_counts().to_dict() # Distribusi nilai target
}

# Menampilkan informasi dasar dataset
data_info # Menampilkan hasil eksplorasi awal
```

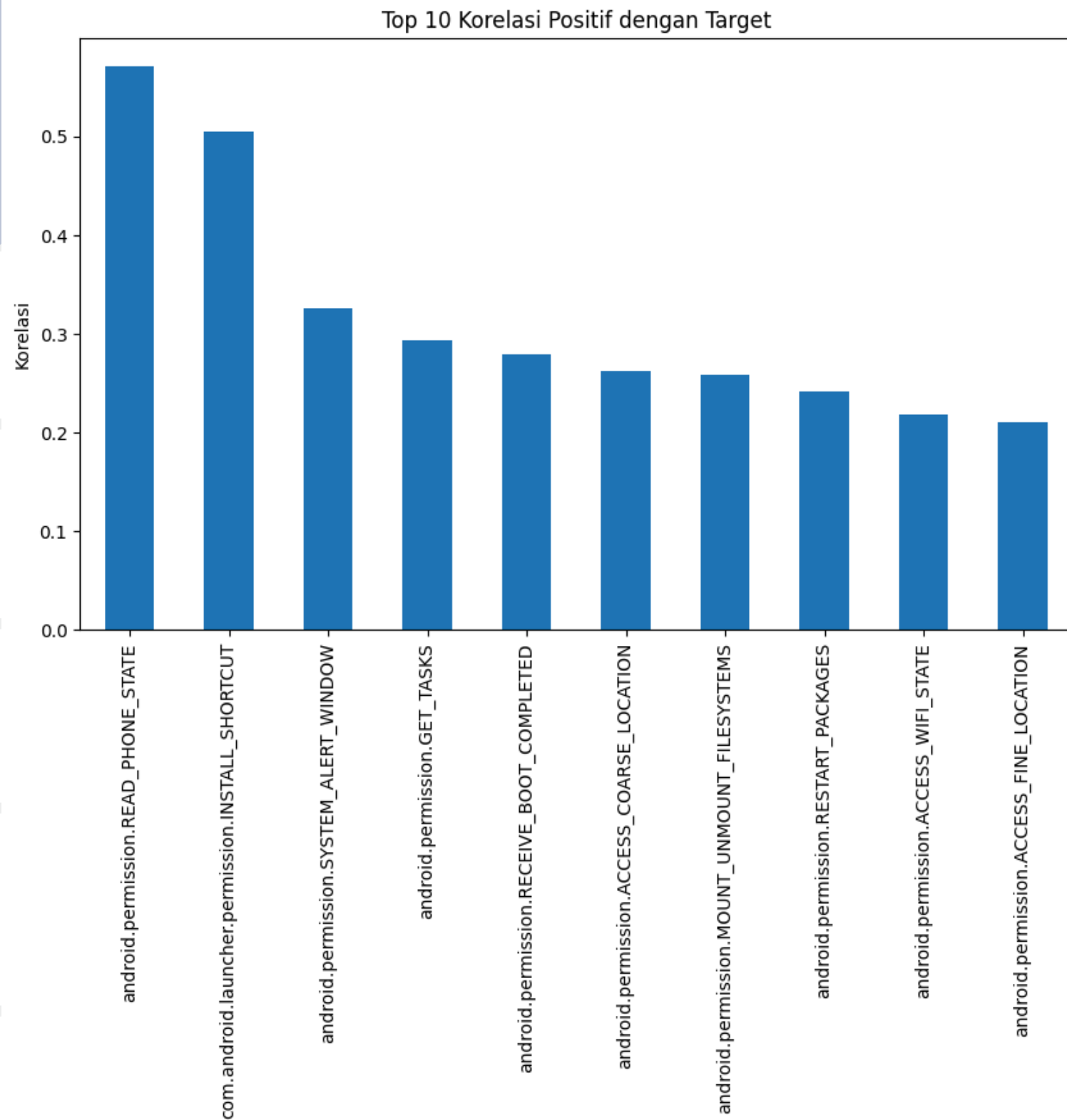
```
{'Shape': (29332, 87),
'Missing Values': 0,
'Duplicate Rows': 21841,
'Target Value Counts': {1: 14700, 0: 14632}}
```

MEMBERIKAN GAMBARAN DASAR TENTANG DATASET, TERMASUK DIMENSI, JUMLAH NILAI YANG HILANG, JUMLAH BARIS DUPLIKAT, DAN DISTRIBUSI NILAI DARI KOLOM TARGET (MISALNYA, 'RESULT').

```
# Menghapus baris duplikat  
data_cleaned = data.drop_duplicates() # Menghapus baris duplikat dari dataset  
  
# Memverifikasi dimensi setelah pembersihan  
data_cleaned.shape # Menampilkan dimensi dataset setelah duplikasi dihapus
```

```
(7491, 87)
```

MENGHAPUS BARIS YANG DUPLIKAT DARI DATASET DAN KEMUDIAN MEMVERIFIKASI DIMENSI DATASET SETELAH PROSES PEMBERSIHAN.



```
import seaborn as sns
import matplotlib.pyplot as plt

# Menghitung matriks korelasi
correlation_matrix = data_cleaned.corr() # Menghitung korelasi antar fitur

# Visualisasi korelasi fitur terhadap target
plt.figure(figsize=(10, 6))
correlation_with_target = correlation_matrix['Result'].sort_values(ascending=False) # Korelasi dengan target
correlation_with_target[1:11].plot(kind='bar', title='Top 10 Korelasi Positif dengan Target') # 10 Korelasi Teratas
plt.ylabel('Korelasi')
plt.show()
```

MENGHITUNG KORELASI ANTAR FITUR DALAM DATASET. LALU MENAMPILKAN GRAFIK BATANG YANG MENUNJUKKAN 10 FITUR TERATAS YANG MEMILIKI KORELASI POSITIF PALING TINGGI DENGAN TARGET (MISALNYA, KOLOM 'RESULT').

```
warnings.warn(smsg, UserWarning)
```

	Model	Accuracy	Best Params
0	Logistic Regression	0.934623	{'classifier__C': 0.1}
1	Decision Tree	0.907939	{'classifier__max_depth': 10, 'classifier__min...
2	k-NN	0.904603	{'classifier__n_neighbors': 5, 'classifier__we...
3	XGBoost	0.941294	{'classifier__max_depth': 5, 'classifier__n_es...



FUNGSI INI MELAKUKAN INISIALISASI BEBERAPA MODEL KLASIFIKASI. KEMUDIAN MELAKUKAN PENCARIAN HYPERPARAMETER TERBAIK MENGGUNAKAN GRIDSEARCHCV DENGAN CROSS-VALIDATION 5 LIPATAN. MODEL YANG DILATIH DIEVALUASI BERDASARKAN AKURASI. DAN HASIL EVALUASI TERSEBUT DISUSUN DALAM BENTUK DATAFRAME YANG MEMUAT MODEL, AKURASI, DAN PARAMETER TERBAIK YANG DITEMUKAN.



TERIMA KASIH