



• • • •

PREDIKSI VOLUME LALU LINTAS DENGAN REGRESI LINEAR DAN POLINOMIAL

KELOMPOK 2





KELOMPOK 2

FARAH NASYWA 2208107010051

IWANI KHAIRINA 2208107010078

DINDA MAHARANI 2208107010081





Pemahaman DATASET

Dataset yang digunakan adalah Metro Interstate Traffic Volume dari UCI Machine Learning Repository.

Dataset ini mencatat volume lalu lintas kendaraan di Interstate 94, Minneapolis, disertai informasi waktu dan cuaca.



Metro Interstate Traffic Volume - UCI.

Data Loading

Langkah ini memuat berbagai library Python yang dibutuhkan, seperti pandas, numpy, matplotlib, dan seaborn untuk manipulasi data dan visualisasi. Selain itu, digunakan juga modul dari sklearn untuk membangun dan mengevaluasi model regresi. Dataset dibaca menggunakan pd.read_csv() dari file compressed CSV, lalu ditampilkan beberapa baris awal menggunakan df.head() untuk melihat isi data secara sekilas.

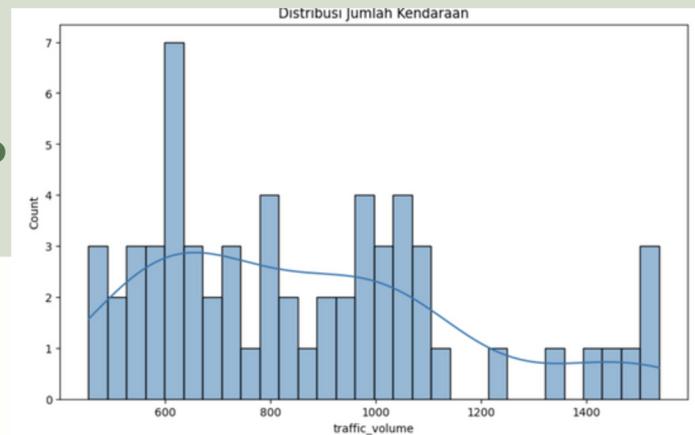
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures, StandardScaler
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error

# Load dataset
df = pd.read_csv('Metro_Interstate_Traffic_Volume.csv.gz')
df.head()
```

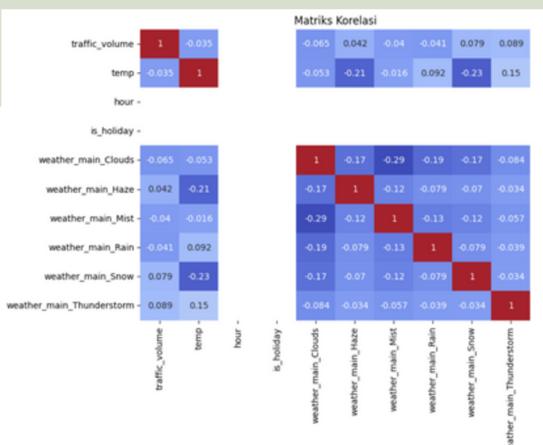


NEXT PAGE 



Eksplorasi dan Pra-pemrosesan

Berdasarkan hasil df.info() dan df.describe(), kolom holiday memiliki banyak nilai kosong sehingga baris tersebut dihapus. Kolom date_time dikonversi ke format waktu dan diambil nilai jamnya sebagai fitur hour. Kolom holiday juga diubah menjadi fitur biner is_holiday. Visualisasi traffic_volume menunjukkan distribusi yang tidak sepenuhnya normal, namun tetap memberikan gambaran umum kepadatan lalu lintas.



Korelasi dan Pemilihan Fitur

Pada tahap ini, dilakukan analisis korelasi antara traffic_volume dan fitur lain seperti temp, hour, is_holiday, serta weather_main (dalam bentuk dummy variables). Heatmap korelasi digunakan untuk mengidentifikasi fitur dengan hubungan terkuat terhadap traffic_volume, yang penting untuk pemilihan fitur dalam model prediksi.



Implementasi Model

• • • •



Linear Regression

MSE: 93207.35125166873

MAE: 276.39050573455876

R2 Score: -0.4670951669749508



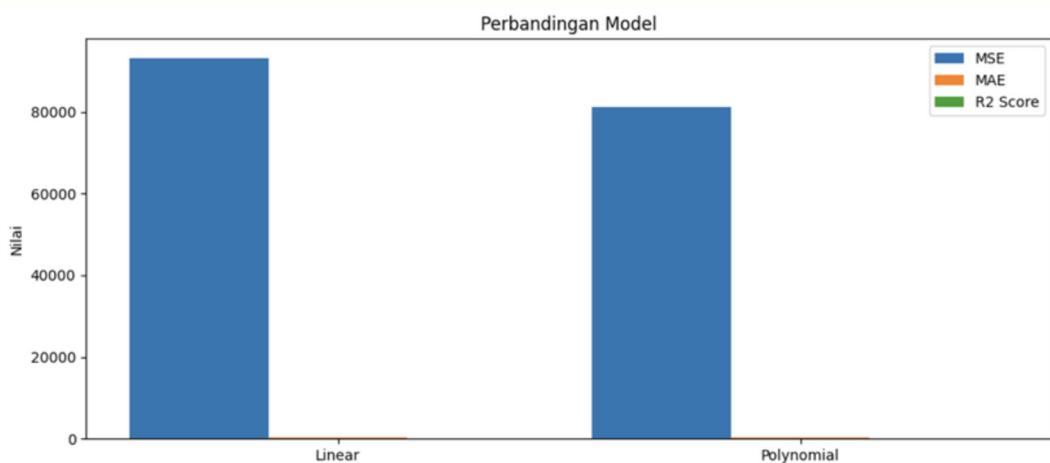
Polynomial Regression

MSE: 81162.0873516974 MAE:

233.1377261433607 R2 Score:

-0.27750123242711755

Perbandingan Model





Interpretasi Hasil

- Model linear memiliki nilai R^2 sebesar -0.467 sedangkan model polinomial memiliki R^2 sebesar -0.277
- Model polinomial menunjukkan hasil evaluasi yang lebih baik namun perlu diperhatikan potensi overfitting.
- Koefisien regresi linear menunjukkan bahwa fitur `jam` dan `kecepatan` berpengaruh signifikan terhadap volume lalu lintas.

kesimpulan

Model regresi polinomial derajat 2 memberikan hasil yang lebih akurat dalam memprediksi jumlah kendaraan per jam dibandingkan model regresi linear biasa.

THANK YOU!

