

<p> PRAKTIKUM DATA MINING </p>

# UJIAN TENGAH SEMESTER

<p> Nicholas Jeremy Hendrajaya </p>  
<p> 2310631170109 </p>



Forms x



# Link Ke Google Colab

- <https://colab.research.google.com/drive/1eai-cuoBdhnWt-0qweqdyu65274VSg12s?usp=sharing>





Forms x

# DATASET

- Dataset: Used Cars Price Prediction
- Sumber: [https://raw.githubusercontent.com/FarrelllAdityaaa/dataset-uts-datamining/refs/heads/main/used\\_cars\\_price\\_fiks.csv](https://raw.githubusercontent.com/FarrelllAdityaaa/dataset-uts-datamining/refs/heads/main/used_cars_price_fiks.csv)

```
[ ] # Load dataset mobil bekas
!wget -q -O used_cars.csv "https://raw.githubusercontent.com/FarrelllAdityaaa/dataset-uts-datamining/refs/heads/main/used_cars_price_fiks.csv"
```

```
[ ] dataset = pd.read_csv("used_cars.csv")
print("Shape dataset:", dataset.shape)
dataset.head()
```

↗ Shape dataset: (6019, 13)

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	Price
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000.0	CNG	Manual	First	26.60	998.0	58.16	5.0	1.75
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000.0	Diesel	Manual	First	19.67	1582.0	126.20	5.0	12.50
2	2	Honda Jazz V	Chennai	2011	46000.0	Petrol	Manual	First	18.20	1199.0	88.70	5.0	4.50
3	3	Maruti Ertiga VDI	Chennai	2012	87000.0	Diesel	Manual	First	20.77	1248.0	88.76	7.0	6.00
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670.0	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	17.74



Forms x



# PREPROCESSING

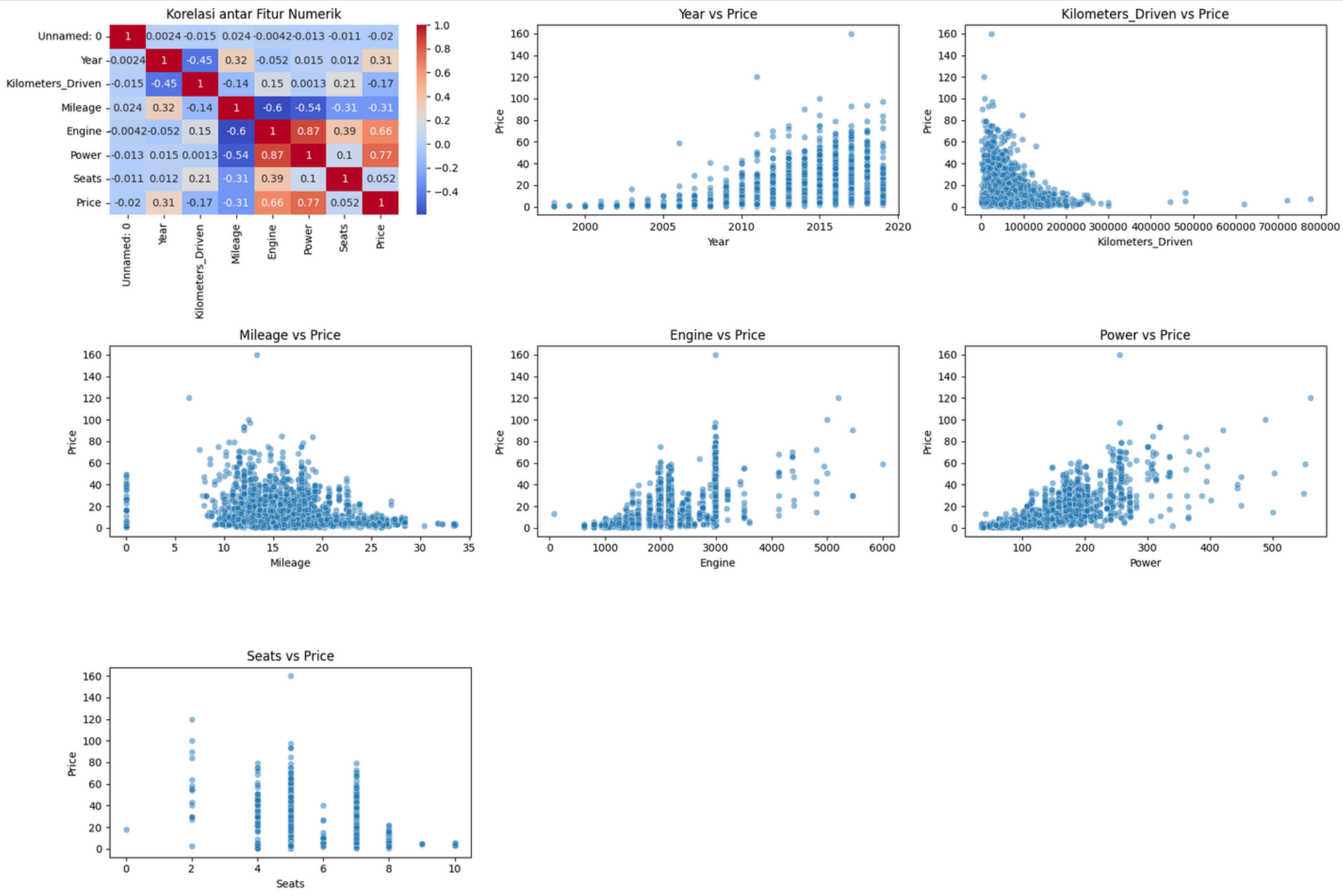
- Profiling Data: melihat tipe data, statistik deskriptif, dan missing value
- Exploratory Data Analysis (EDA): histogram harga, boxplot outlier, grafik batang kategorikal, heatmap korelasi
- Data Cleaning: hapus nilai kosong, duplikat, kolom ID, dan outlier harga
- Data Final: dataset bersih tanpa missing value, siap modelling





Forms ✕

# PREPROCESSING





Forms x

# MODELLING

- Split data: 80% train, 20% test
- Scaling data numerik dengan StandardScaler
- Model yang digunakan: Linear Regression
- Proses: fit model → prediksi harga mobil

```
[ ] # Membangun model Linear Regression
model = LinearRegression()
model.fit(X_train_final, y_train)

[ ] # Prediksi data uji
y_pred = model.predict(X_test_final)

print("\nContoh hasil prediksi harga mobil:")
print(y_pred[:5])

Contoh hasil prediksi harga mobil:
[ 3.14020426  4.81299799  6.01037749  7.10897196 -0.44592769]
```



Forms x



# EVALUASI

- Evaluasi dengan membandingkan hasil prediksi dengan data aktual dengan metrik:  $R^2$ , MAE, MSE, RMSE
- Interpretasi Metrik:
  - $R^2 \rightarrow$  mendekati 1 berarti model makin baik
  - MAE  $\rightarrow$  rata-rata selisih absolut (semakin kecil semakin baik)
  - MSE  $\rightarrow$  rata-rata selisih kuadrat (memberi penalti besar ke error besar)
  - RMSE  $\rightarrow$  akar MSE, mudah dipahami karena sama satuannya dengan harga mobil





Forms x



# EVALUASI

• Berikut adalah hasil dari evaluasi model:

[ ]

```
# Hitung metrik evaluasi
r2_test = r2_score(y_test, y_pred_test)
mae_test = mean_absolute_error(y_test, y_pred_test)
mse_test = mean_squared_error(y_test, y_pred_test)
rmse_test = math.sqrt(mse_test)

# Buat DataFrame nilai evaluasi
score = pd.DataFrame({
    'Metrik': ['R²', 'MAE', 'MSE', 'RMSE'],
    'Nilai': [r2_test, mae_test, mse_test, rmse_test]
})

print("\nHasil evaluasi model:")
display(score)
```



Hasil evaluasi model:

	Metrik	Nilai
0	R²	0.699125
1	MAE	1.592233
2	MSE	5.078290
3	RMSE	2.253506



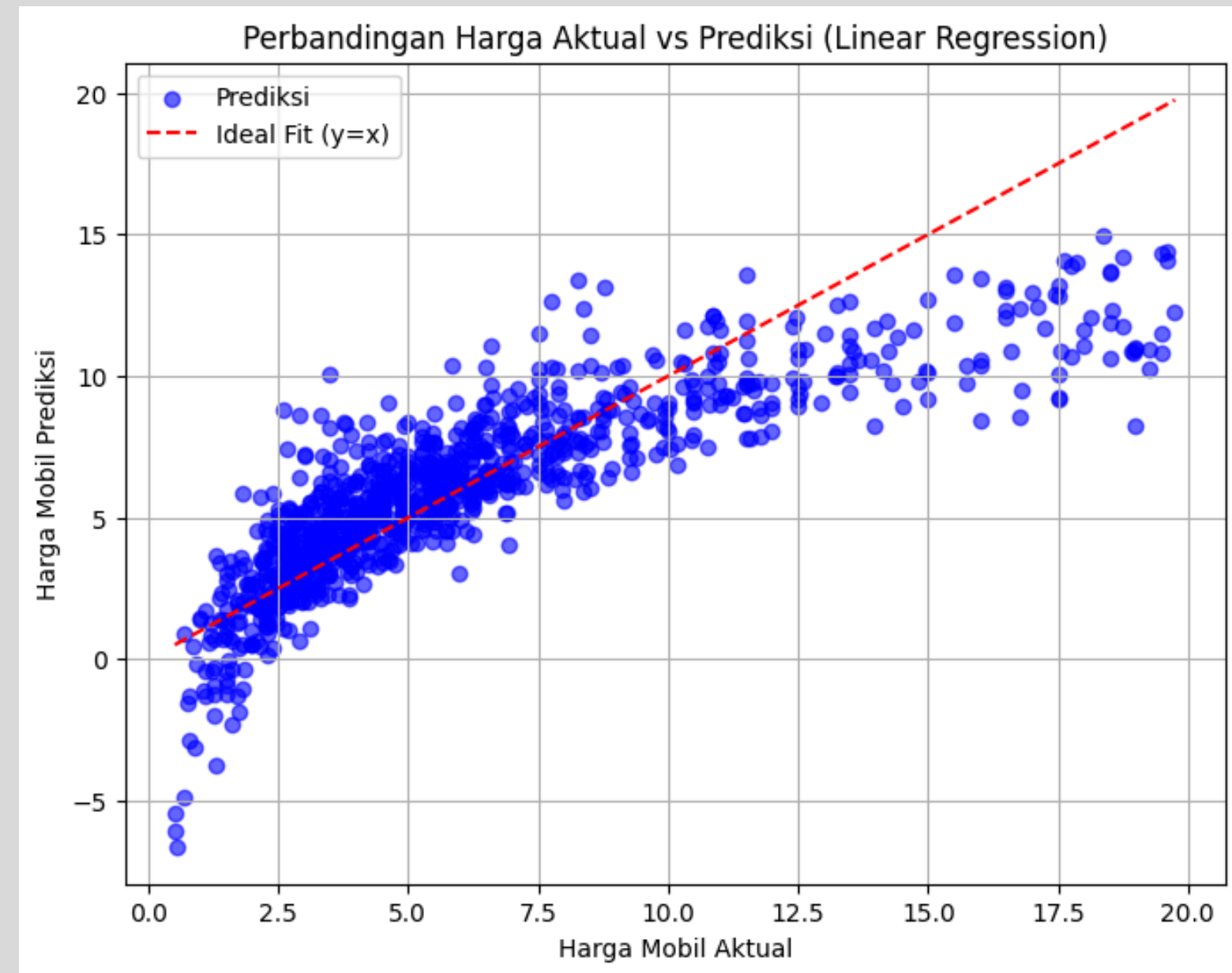




Forms x

# EVALUASI

- Berikut adalah hasil nilai prediksi dan nilai aktual dari model dan dataset tersebut:





Forms

Input x



## KESIMPULAN

Pada penelitian ini digunakan dataset mobil bekas dengan 6.019 baris dan 13 kolom, dengan target prediksi berupa harga mobil. Setelah melalui tahap preprocessing yang meliputi profiling, eksplorasi data, pembersihan data, dan menghasilkan data final, model Linear Regression dibangun dengan melakukan scaling pada data numerik.

Hasil evaluasi menunjukkan nilai  $R^2$  sebesar 0,699 yang berarti model cukup baik dalam menjelaskan variasi harga mobil, dengan tingkat error yang relatif kecil berdasarkan MAE, MSE, dan RMSE. Dari sini dapat disimpulkan bahwa model Linear Regression mampu memprediksi harga mobil bekas dengan cukup akurat, meskipun masih ada ruang perbaikan, misalnya dengan mencoba algoritma lain atau menambahkan fitur tambahan.



**<p> THANK YOU  
FOR LISTENING </p>**

**Nicholas Jeremy Hendrajaya - 2310631170109**