Universitas
Singaperbangsa
Karawang

PREDIKSI HARGA MOBIL BEKAS MENGGUNAKAN LINEAR REGRESSION

Prepared by
Salsa Anderia Putri
Nabila

Universitas
Singaperbangsa
Karawang

LATAR BELAKANG

- Harga mobil bekas sangat bervariasi tergantung faktor seperti tahun, kilometer tempuh, dan kapasitas mesin.
- Diperlukan model prediktif untuk memperkirakan harga mobil bekas secara lebih akurat.
- Proyek ini menggunakan metode Linear Regression untuk memprediksi harga berdasarkan dataset dari Kaggle.

DATASET

Sumber datasetnya:
- *Kaggle - Used Cars Price Prediction*
- *Github - User Cars Price Prediction*

Jumlah data: 6.019 baris dan 25 kolom

Fitur penting:
- *Year, Kilometers_Driven, Mileage, Engine, Power, Seats*
- *Kategori seperti: Location, Fuel_Type, Transmission, Owner_Type*

Target: Price (harga mobil bekas)

Universitas
Singaperbangsa
Karawang

1.PROFILING DATA: MENGETAHUI TIPE DATA DAN JUMLAH NILAI KOSONG.

[5]
✓ 0s

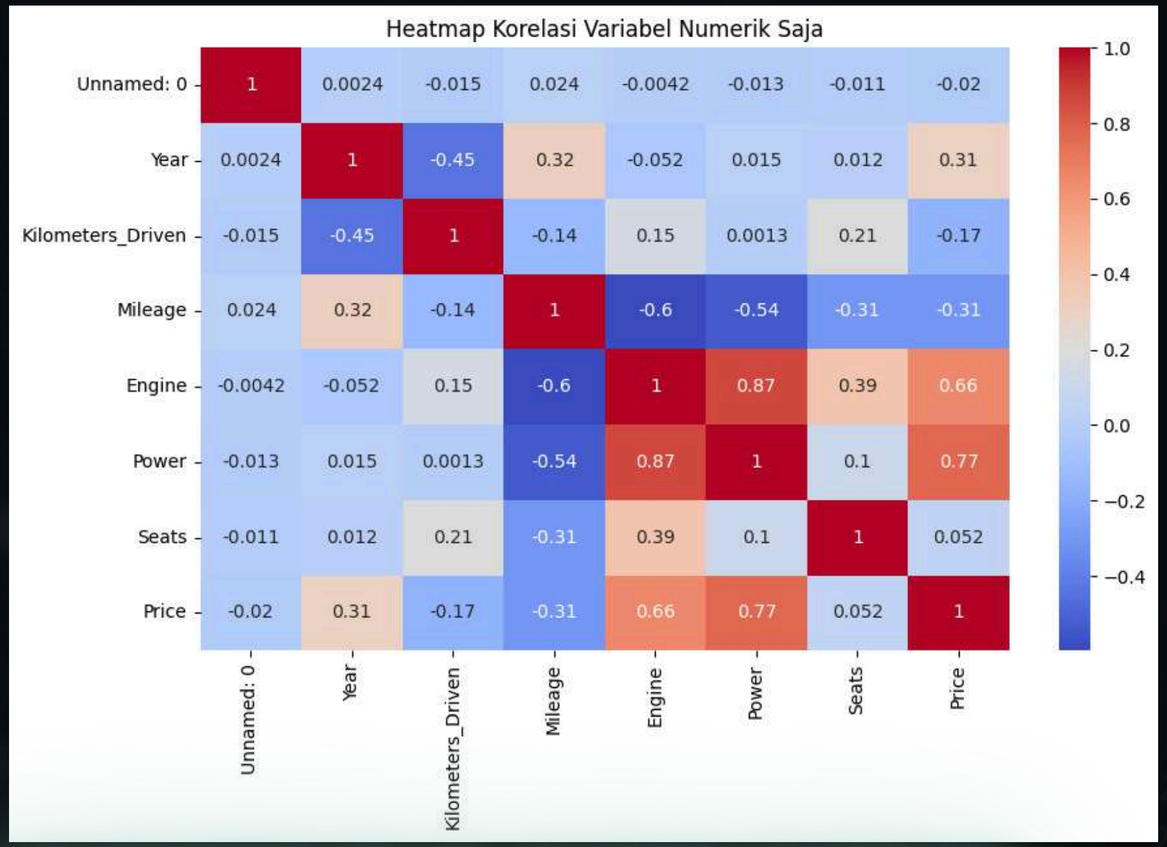
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6019 entries, 0 to 6018
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Unnamed: 0         6019 non-null   int64
 1   Name               6019 non-null   object
 2   Location           6019 non-null   object
 3   Year               6019 non-null   int64
 4   Kilometers_Driven  5719 non-null   float64
 5   Fuel_Type          6019 non-null   object
 6   Transmission       6019 non-null   object
 7   Owner_Type         6019 non-null   object
 8   Mileage            6017 non-null   float64
 9   Engine             5983 non-null   float64
 10  Power              5876 non-null   float64
 11  Seats              5977 non-null   float64
 12  Price              6019 non-null   float64
dtypes: float64(6), int64(2), object(5)
memory usage: 611.4+ KB
```

TAHAP PREPROCESSING

Universitas
Singaperbangsa
Karawang

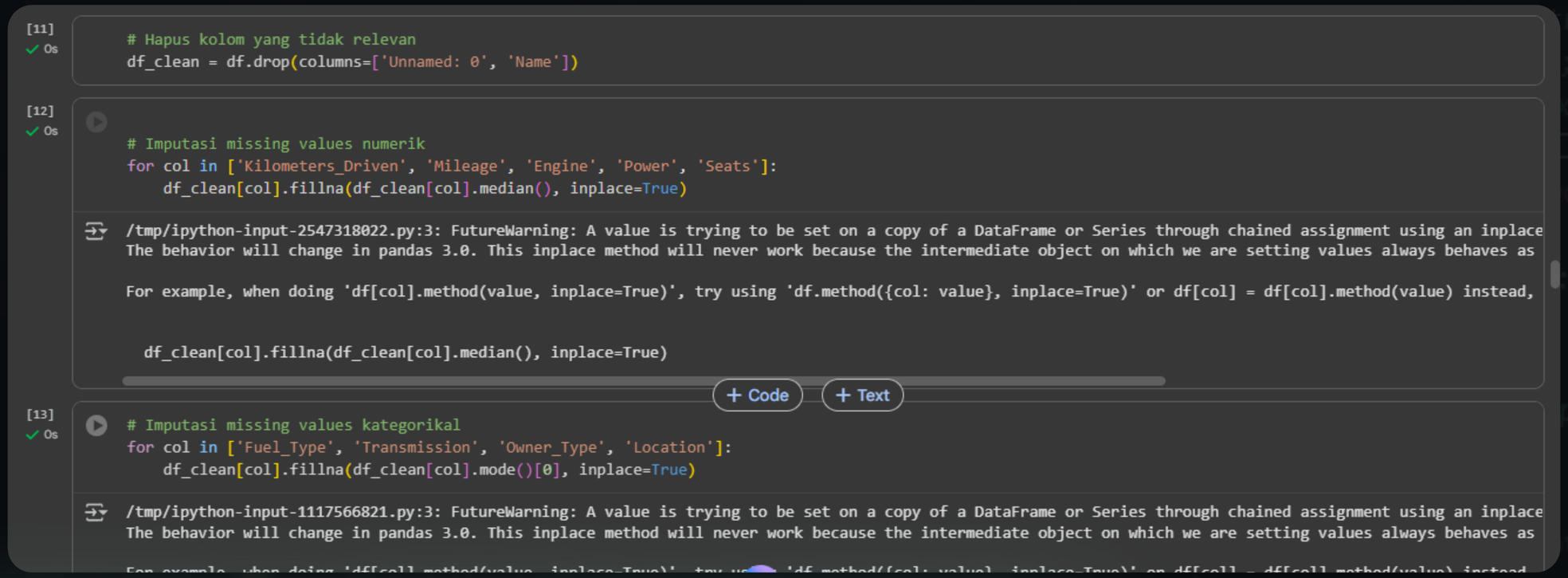
2. EXPLORATORY DATA ANALYSIS (EDA): MELIHAT SEBARAN DATA DAN KORELASI ANTAR FITUR.



TAHAP PREPROCESSING

Universitas Singaperbangsa Karawang

3. DATA CLEANING: MENGHAPUS ATAU MENGISI MISSING VALUES DAN MENGATASI OUTLIER.



[11] ✓ 0s

```
# Hapus kolom yang tidak relevan
df_clean = df.drop(columns=['Unnamed: 0', 'Name'])
```

[12] ✓ 0s

```
# Imputasi missing values numerik
for col in ['Kilometers_Driven', 'Mileage', 'Engine', 'Power', 'Seats']:
    df_clean[col].fillna(df_clean[col].median(), inplace=True)
```

```
/tmp/ipython-input-2547318022.py:3: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead,


  df_clean[col].fillna(df_clean[col].median(), inplace=True)
```

+ Code + Text

[13] ✓ 0s

```
# Imputasi missing values kategorikal
for col in ['Fuel_Type', 'Transmission', 'Owner_Type', 'Location']:
    df_clean[col].fillna(df_clean[col].mode()[0], inplace=True)
```

```
/tmp/ipython-input-1117566821.py:3: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as
```

TAHAP PREPROCESSING

Universitas Singaperbangsa Karawang

3. DATA CLEANING: MENGHAPUS ATAU MENGISI MISSING VALUES DAN MENGATASI OUTLIER.

```
df_clean[col].fillna(df_clean[col].mode()[0], inplace=True)
```

[14]
✓ 0s

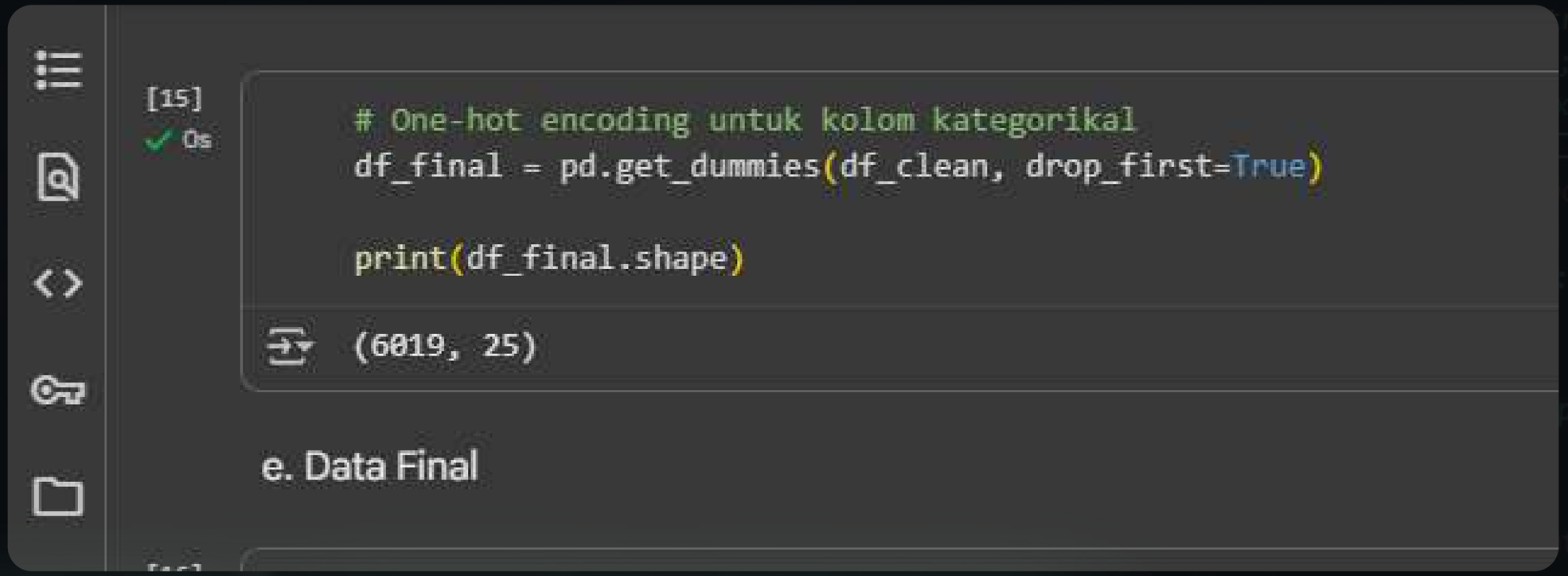
```
# Cek ulang
print(df_clean.isna().sum())
```

```
Location             0
Year                 0
Kilometers_Driven    0
Fuel_Type            0
Transmission         0
Owner_Type           0
Mileage              0
Engine               0
Power                0
Seats                0
Price                0
dtype: int64
```

TAHAP PREPROCESSING

Universitas
Singaperbangsa
Karawang

4. TRANSFORMASI DATA:MENGUBAH DATA KATEGORI MENJADI VARIABEL DUMMY (ONE -HOT ENCODING).



TAHAP PREPROCESSING

Universitas
Singaperbangsa
Karawang

5. DATA FINAL: DATASET SIAP UNTUK MODELLING, SEMUA NILAI NUMERIK DAN TIDAK ADA MISSING VALUE.

```
print(df_final.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6019 entries, 0 to 6018
Data columns (total 25 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Year                      6019 non-null   int64
 1   Kilometers_Driven         6019 non-null   float64
 2   Mileage                   6019 non-null   float64
 3   Engine                    6019 non-null   float64
 4   Power                     6019 non-null   float64
 5   Seats                     6019 non-null   float64
 6   Price                     6019 non-null   float64
 7   Location_Bangalore        6019 non-null   bool
 8   Location_Chennai          6019 non-null   bool
 9   Location_Coimbatore       6019 non-null   bool
 10  Location_Delhi            6019 non-null   bool
 11  Location_Hyderabad        6019 non-null   bool
 12  Location_Jaipur           6019 non-null   bool
 13  Location_Kochi            6019 non-null   bool
 14  Location_Kolkata          6019 non-null   bool
 15  Location_Mumbai           6019 non-null   bool
 16  Location_Pune             6019 non-null   bool
 17  Fuel_Type_Diesel          6019 non-null   bool
 18  Fuel_Type_Electric        6019 non-null   bool
```

TAHAP PREPROCESSING

Universitas Singaperbangsa Karawang

5. DATA FINAL: DATASET SIAP UNTUK MODELLING, SEMUA NILAI NUMERIK DAN TIDAK ADA MISSING VALUE.

```
 18  Fuel_Type_Electric          6019 non-null   bool
 19  Fuel_Type_LPG               6019 non-null   bool
 20  Fuel_Type_Petrol            6019 non-null   bool
 21  Transmission_Manual         6019 non-null   bool
 22  Owner_Type_Fourth & Above   6019 non-null   bool
 23  Owner_Type_Second           6019 non-null   bool
 24  Owner_Type_Third            6019 non-null   bool
dtypes: bool(18), float64(6), int64(1)
memory usage: 435.1 KB
None
```

TAHAP PREPROCESSING

Universitas Singaperbangsa Karawang

MODELLING

3. Melakukan Data Modeling

[22]
✓ 0s

```
# Pisahkan fitur dan target
X = df_final.drop(columns=['Price'])
y = df_final['Price']

# Train-test split (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# Data Scaling (StandardScaler)
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

+ Code + Text

Universitas
Singaperbangsa
Karawang

MODELLING

[31]
✓ 0s

```
# Model Linear Regression
model = LinearRegression()
model.fit(X_train_scaled, y_train)
```

```
LinearRegression
LinearRegression()
```

[29]
✓ 0s

```
# Prediksi
y_pred = model.predict(X_test_scaled)
```

+ Code

Universitas
Singaperbangsa
Karawang

EVALUASI MODEL

4. Melakukan tahap evaluasi model

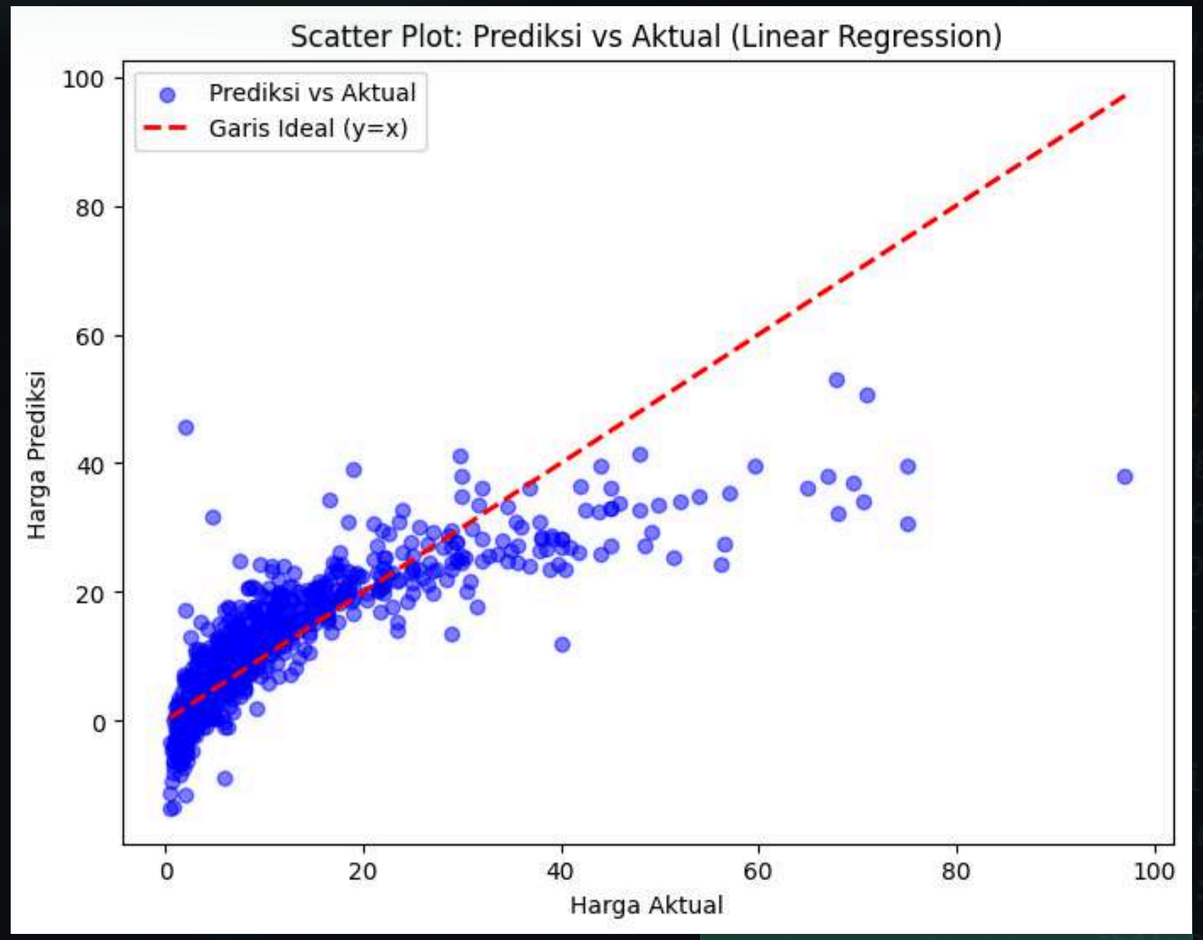
```
# Evaluasi model
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
```

```
print("Evaluasi Model Linear Regression")
print(f"R-squared (R²): {r2:.4f}")
print(f"Mean Absolute Error (MAE): {mae:.4f}")
print(f"Mean Squared Error (MSE): {mse:.4f}")
print(f"Root Mean Squared Error (RMSE): {rmse:.4f}")
```

```
Evaluasi Model Linear Regression
R-squared (R²): 0.6930
Mean Absolute Error (MAE): 3.7981
Mean Squared Error (MSE): 37.7749
Root Mean Squared Error (RMSE): 6.1461
```

Universitas
Singaperbangsa
Karawang

VISUALISASI HASIL



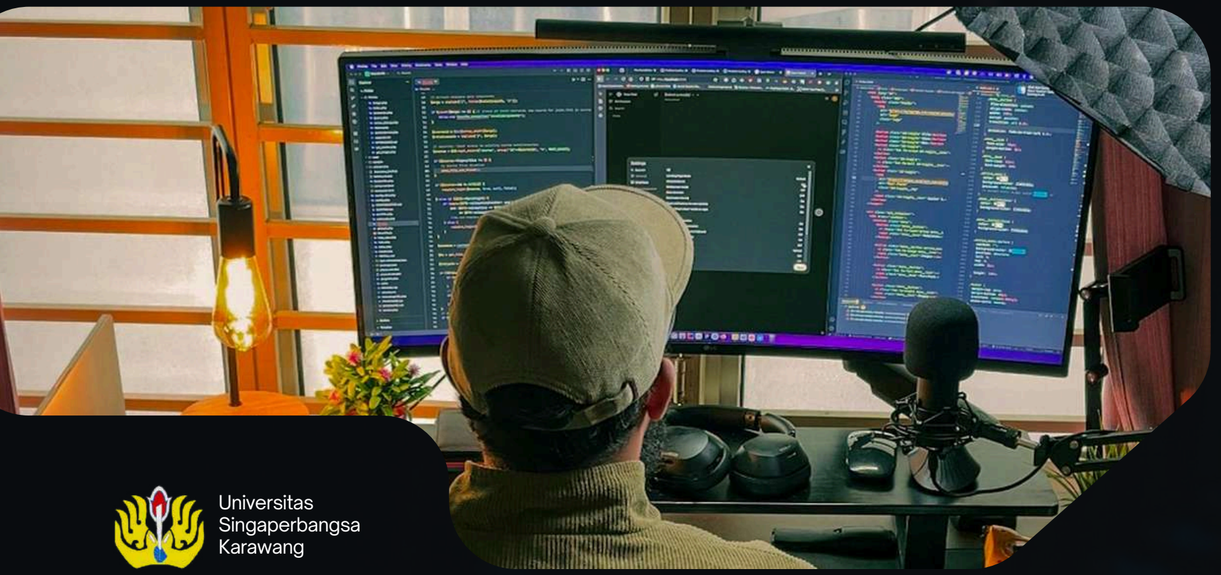
Universitas
Singaperbangsa
Karawang

Universitas
Singaperbangsa
Karawang

- Model Linear Regression berhasil memprediksi harga mobil bekas dengan tingkat akurasi yang cukup baik ($R^2 = 0.69$).
- Fitur seperti tahun kendaraan, kilometer tempuh, daya mesin, dan jenis bahan bakar berpengaruh besar terhadap harga.
- Untuk hasil yang lebih optimal, dapat dicoba model lain seperti Random Forest atau Gradient Boosting.

KESIMPULAN

THANK YOU



Universitas
Singaperbangsa
Karawang

REFERENSI :

- Sumber pembelajaran: Praktikum Data Mining

Sumber datasetnya:

- Kaggle - Used Cars Price Prediction
- Github - User Cars Price Prediction