

TUGAS 1
PERSOALAN REGRESI
PEMBELAJARAN MESIN LANJUT

Oleh
MOHAMAD IDRIS
NIM : 33221010



SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
TAHUN 2022

Tugas 1

Persoalan Regresi

dengan Fully Connected Neural Network dan dataset Boston Housing Price

Menggunakan Numpy dan Pandas untuk memproses dataset, matplotlib dan seaborn untuk melakukan visualisasi pada data, serta Keras untuk mengimplementasikan jaringan saraf (Neural Network) seperti yang terlihat pada Gambar 1 berikut:

```
!pip install keras-tuner --upgrade
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
from sklearn.preprocessing import StandardScaler # Standardization
from sklearn.ensemble import IsolationForest # Outlier Detection
from keras.models import Sequential # Sequential Neural Network
from keras.layers import Dense
from keras.callbacks import EarlyStopping # Early Stopping Callback
from tensorflow.keras.optimizers import Adam # Optimizer
from kerastuner.tuners import RandomSearch # HyperParameter Tuning
from sklearn.preprocessing import StandardScaler
import warnings
warnings.filterwarnings('ignore') # To ignore warnings.
```

Gambar 1. Proses Import Library

Proses Selanjutnya adalah memuat dataset menggunakan Boston Housing Price, seperti yang diperlihatkan pada Gambar 2 berikut:

```
train = pd.read_csv('house-prices-advanced-regression-techniques/train.csv')
test = pd.read_csv('house-prices-advanced-regression-techniques/test.csv')
y = train['SalePrice'].values
data = pd.concat([train, test], axis=0, sort=False)
data.drop(['SalePrice'], axis=1, inplace=True)
data.head()
```

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	...
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	...
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	...
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	...

5 rows × 80 columns

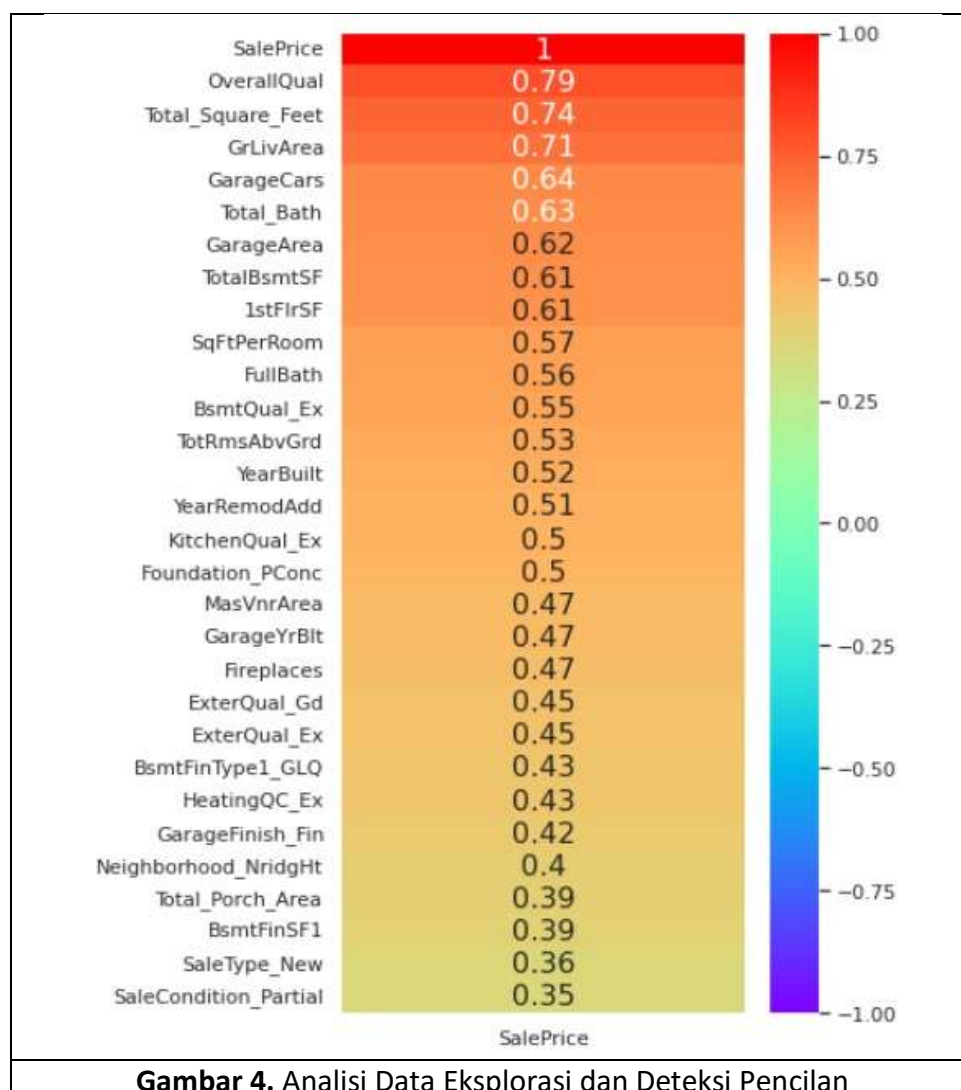
Gambar 2. Pembagian Dataset

Analisis dan Imputasi Nilai yang Hilang, yaitu melihat secara keseluruhan fitur apa saja yang nilainya hilang. Dalam kasus ini terdapat sekitar 33 fitur hilang yang diuraikan secara lengkap seperti terlihat pada Gambar 3 berikut:

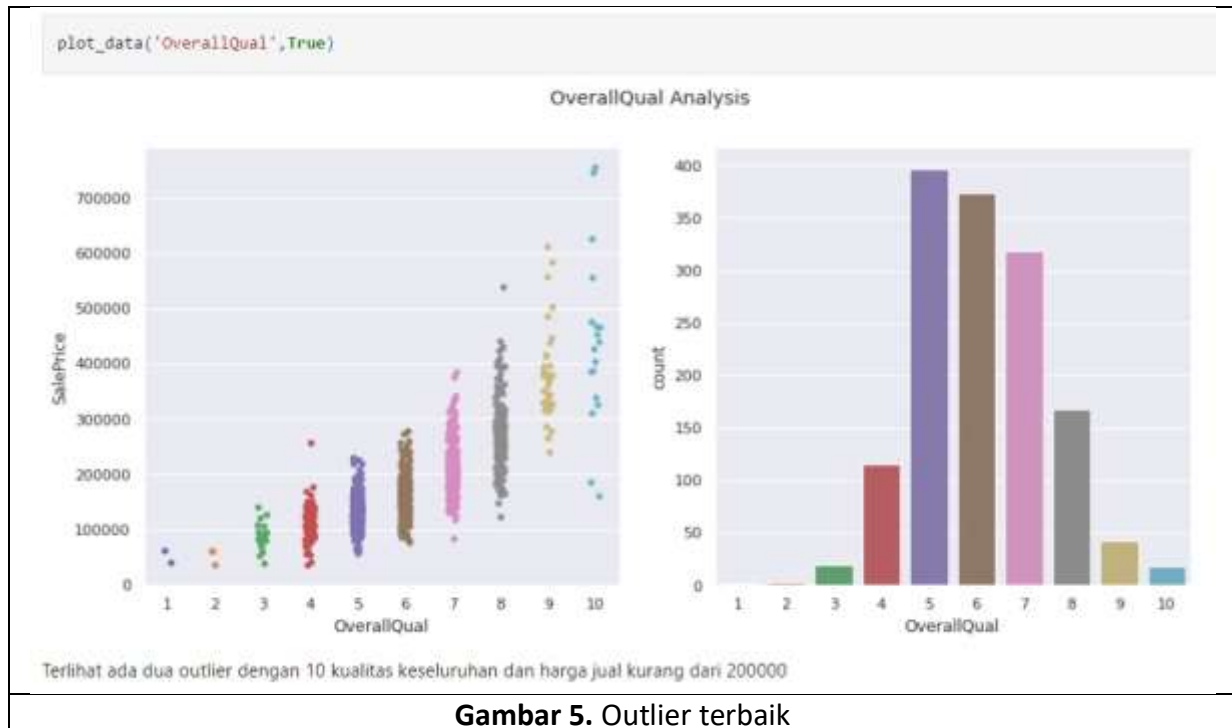
	Feature	Number of Missing Values	Percentage of Missing Values
0	PoolQC	2909	99.657417
1	MiscFeature	2814	96.402878
2	Alley	2721	93.216855
3	Fence	2348	80.438506
⋮			
31	Exterior2nd	1	0.034258
32	Exterior1st	1	0.034258
33	SaleType	1	0.034258

Gambar 3. Fitur yang hilang

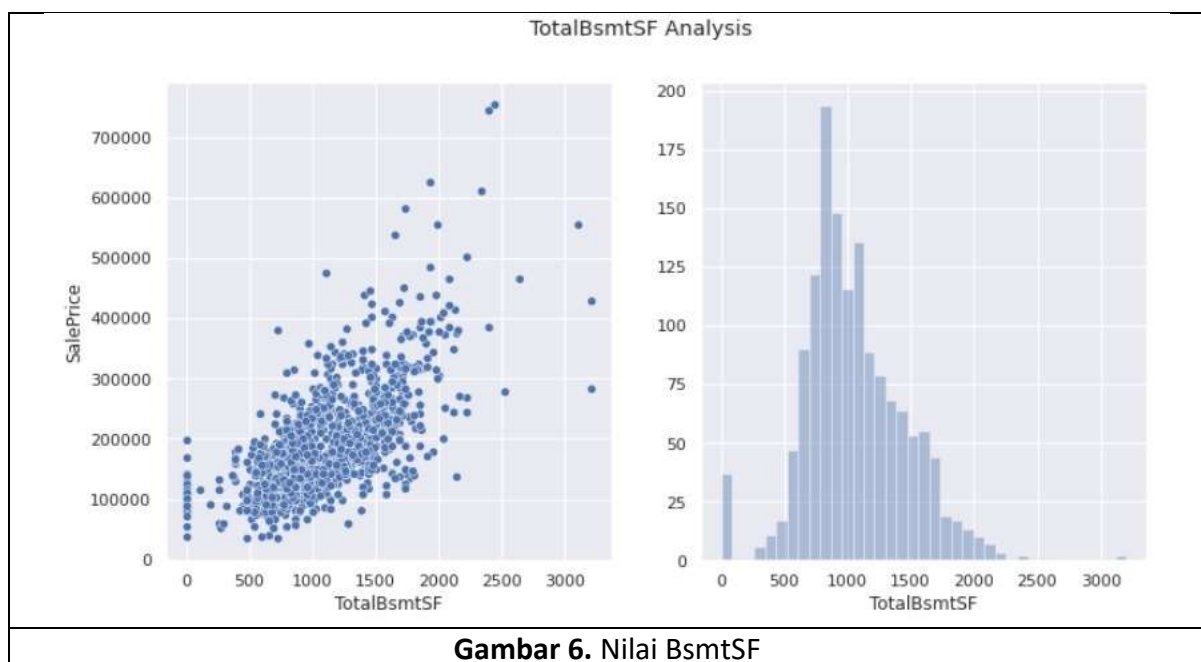
Analisi Data Eksplorasi dan Deteksi Pencilan, yang bertujuan untuk menemukan nilai terbaik dari optimizer, Pertama-tama kita akan mengekstrak fitur teratas dari kumpulan data pelatihan yang memiliki korelasi tertinggi dengan harga jual seperti yang terlihat pada Gambar 4 berikut:



Outlier terbaik yang diperoleh dapat dilihat pada Gambar 5 berikut:



Nilai Total BsmtSF dapat diamati dari Gambar 6 berikut:



Hasil Prediksi dan Evaluasi berdasarkan Fully Connected Neural Network bisa dilihat sebagai berikut:

```
X_test = scale.transform(test) # Scaling the testing data.  
result = model.predict(X_test) # Prediction using model  
result = pd.DataFrame(result, columns=['SalePrice']) # Dataframe  
result['Id'] = test['Id'] # Adding ID to our result dataframe  
result = result[['Id', 'SalePrice']]  
result.head()
```

	Id	SalePrice
0	1461	51.339394
1	1462	35.059734
2	1463	2001.645142
3	1464	3327.788330
4	1465	2786.517334

Gambar 7. Hasil Prediksi dan Evaluasi