Advanced Predict House Price

Ari Sulistyowati

Steps for an Advanced Predict House Price

1. Overview Dataset
2. Data Preparation
3. Exploratory Data Analysis
4. Split Dataset
5. Built Machine Learning
6. Best Model
7. Evaluation Model

Step 1: Overview Dataset

In this stage, I aplicated to import library python to helping me built machine learning models. First, I am import library such as:

a. pandas,

b. numpy,

c. matplotlib,

d. seaborn,

e. warning filter,

f. sklearn.preposessing,

g. scipy.stats.mstats import winsorize

h. sklearn.linear_model, sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error

Step 2: Data Preparation

- First, I calculated correlation between feature column and target column. From this step, several columns obtained have a correlation with target column more than 0.5 poin. For this reason, this columns will be further processed.

- Second, I divide it into two stages namely, checking missing values and checking outliers.

1. Checking Missing Values

   a) I created a function to count all the columns containing missing values in dataset by dividing one hundred so that I get the percent value.

   b) A deleted column has missing values percent higher 80%.

   c) Treatment of missing values column which has a value below 80 percent by using mean.

Step 2: Data Preparation

2. Checking Outliers

   In this step I a just using winsorization method:

   a) Counting outer fence outliers.
   b) Visualization outer fence outliers with data column.
   c) Chossing the right quantile.
   d) Counting to choose quantile.
   e) Choose quantile.

Step 3: Exploratory Data Analysis

At this stage, I use visualization to understand more about the dataset. First, I divided into several categories:

a. Analysis of Dataset.

  To understand about basic data in dataset.

b. Analyze feature column with target column.

  To see linearity o the features column and the target column.

c. Analysis of Year and SalePrice.

d. Analysis of Area with SalePrice.

e. Deskriptive Statistik.

Step 4: Split Dataset

- Split the data into training data and test data with *train test split*.
- test_size=0.2, random_state=4

Step 5: Built Machine Learning

- For making machine learning models, I use various kinds of linear regression and regularization models.
- Machine learning model I used such as:

a)   Multiple Linear Regression

b)   Ridge

c)   Lasso

d)   Elastic Net

Step 6: Best Model

- For the selection model I rate on the highest score accuracy.
- Score accuracy is an assessment of how much percent machine learning modeling recognizes data in a particular dataset.

Step 7: Evaluation Model

- At this stage I calculate the accuracy of the model from several aspects based on:

a) R2_Score

b) MAE_Score

c) MSE_Score

Thank You

Ari Sulistyowati