Housing Price Prediction

Submitted by:

Anita Tarangi

Introduction

* Business Problem Framing

You are required to model the price of houses with the available independent variables. This model will then be used by the management to understand how exactly the prices vary with the variables. They can accordingly manipulate the strategy of the firm and concentrate on areas that will yield high returns. Further, the model will be a good way for the management to understand the pricing dynamics of a new market.

Analytical Problem Framing

* Data Sources and their formats

Dataset is in CSV format with many null values.

* Data Pre-processing Done

1. **Checking the size of the dataset**
2. **Checking the summary statistics of the dataset**
3. **Checking the data types and null values**
4. **Dropping NULL’S for columns above 15%**
5. **Performing an Imputation for columns having null values below 15%**
6. **Dropping columns which aren’t important**

* Hardware and Software Requirements and Tools Used

Python, Pandas, Seaborn, Matplotlib, sklearn, missingno,statsmodel

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches Regression methods are used, and Lasso Regression performed better than other models.
* Testing of Identified Approaches

Recursive feature elimination (RFE), VIF, Residual Analysis

* Run and evaluate selected models

Linear Regression, Ridge Regression, Lasso Regression

* Visualizations

Seaborn, Matplotlib, Missingno

* Interpretation of the Results

1. The R2 Score, RSS and MSE are all very close for Linear Regression, Ridge and Lasso
2. Lasso has better scores by a very slight margin compared to Ridge and Linear Regression

**CONCLUSION**

* Comparing with Linear and Ridge Regression, Lasso Regression performed slightly better.