Capstone Project Proposal

House Prices: Advanced Regression Techniques

Problem

Prediction of sale prices.

Let's say that a buyer is interested in purchasing a house. He has an estimate of the price of the house and has an offer in mind. The price estimation might have been based on few factors or external sources such as real estate agencies. The problem for the buyer is knowing the exact amount for the purchase price of the house. For a real estate company, which can also pose as a buyer or broker, the problem is to negotiate for the best deal. This dataset has several factors. It becomes crucial to know the levers that drive the price and develop a model to predict them with best accuracy.

Client

Model for price prediction of a house can be a valuable tool for buyer/seller of a house, real estate agent/company, builders or tax departments.

Uses of the price prediction model

1. The client will be able to predict sale price of a house.

2. Various aspects or features that have a strong influence on price can be known.

3. The client can be in an advantageous position while negotiating.

4. The model can be useful to real estate agents and online companies as it would save additional costs and time in further examination and research.

5. Having an idea of the most influential features would enable the client to plan and effect changes in the property vis a vis the cost and expected return from investment. One can

also decide what features need to be included for the house construction / renovation as per budget.

Deliverables

The analysis will be done using python program in Jupyter notebook. A final presentation for the same will prepared to highlight the findings.

Data acquisition

The Ames Housing dataset part of Kaggle competitions is used. The dataset has 79 explanatory variables describing every aspect of residential homes in Ames, Iowa.

Process

1. Data cleaning to handle Nan values and outliers.

2. Data observation to understand the scope challenges, estimators data types and central tendencies.

3. Data visualisation to look at the distribution and other aspects of features and target variable.

4. Exploratory Data Analysis to understand the correlations of features among itself and with the target variable.

5. Fit the Linear Model and predict prices.

Metric

Model will be evaluated on Root mean squared error(RMSE) between logarithm of predicted value and logarithm of sale price.