Preliminary Course Report

EAS 503: Programming and Database Fundamentals for Data Scientists

Group Number : 14
Team Members:

Name	UBIT Name	UBID
Sanjana Chalavadi	schalava	50442673
Sai Chetan Thalla	saicheta	50411058
Kranthi Kumar Thummalacherla	kthummal	50432888

Introduction:

Currently, most customers depend on real estate agents to purchase a house. Unfortunately the real estate agents are not fair at all times. They charge hefty premiums and sometimes even inflate the actual price of the property. The aim of our project is to bring fairness in the real estate sectors by providing the customers with the right tools to estimate the real price that the property is valued at. On most occasions much of the data is available on the cloud in some or other forms. The customers just lack the right software to perform the calculations required in price estimation. Not just the customers, even the real estate firms need better models of prediction to be fair and profitable. We will be using standard machine learning techniques like bagging and boosting models to make the predictions.

Problem Statement:

A property company wants to predict the correct price of the properties they have in their database, since people complain about the over inflated prices that the real estate agents sell properties at. The agents only account for a few parameters such as number of rooms, carpet area, locality, commute time etc. The price of any property is composed of two components, the first is demand-supply dynamics, the latter is the inherent value. Inherent value of a house is the function of mostly physical attributes such as style, size, make, garage, locality and zoning, width of street width, noise, lot size and configuration, topography, proximity to key locations, age and finish of the dwelling etc. involved in the sale. The precise problem statement is to make use of the extensive set of data parameters available and make the most accurate prediction as possible.

Dataset description:

We will be using the Ames housing dataset compiled by Dean De Cock. The data set comprises over 80 features that focus on the various elements. There are separate tables for house attributes, extensive attributes, sales records among others. We will be aggregating these data sets in order to build prediction models.

References

1. http://jse.amstat.org/v19n3/decock.pdf