**House Price Prediction**

**Abstract:**

The real estate market is one of the fields where machine learning can be utilized to optimize and predict price with high accuracy. Determining house pricing is an important model for clients to use when making decisions because it uses various parameters to predict the price of a desired property. Many analytical techniques are available to estimate the property price based on a range of characteristics such as the surrounding environment, other facilities, and so on, but the participants in the process are unaware of them. Because of the design, customers will be able to invest in a property without having to go via an agent. Because of its convertible and probabilistic model selection process, the Random Forest model is adopted. The result demonstrates that in order to make comparable forecasts, the solution to the problem must be successful.

**Introduction:**

The real estate industry is a large industry with several stakeholders. There is a need for a greater grasp of the industry's operational strategies and driving variables among various stakeholders. Our project's purpose is to determine the true market worth of the property. This will assist customers in obtaining an accurate estimated price for a certain property. It also provides a pricing study of a property in a certain location.

**Model:**

We made use of Random Forest and Gradient Boosting techniques to predict the prices of houses. Random Forest is a supervised learning algorithm that is used to solve regression and classification problems which uses ensemble learning methods. Gradient Boosting, like Random Forest, is used to solve regression and classification tasks. The difference between the two is that Random Forest combines a large number of trees at the end of the process whereas Gradient Boosting combines trees at the beginning of the process. Random Forest gave a better prediction for our dataset.

**Data:**

For this project, we will be using Ames housing dataset compiled by Dean De Cock. The data set comprises 80 features which focus on the quality and quantity of many physical properties of the property and 3000 observations of the various properties collected from various sources.

**Analysis and Results:**

We included the following main steps: Exploratory Data Analysis-this was done to get a better understanding of data which would be helpful later while building our model. Feature Selection- some of the unimportant features that gave redundant information were discarded to avoid overfitting issues. Modeling- We tried different Machine Learning models to train our data set such as Random Forest and Gradient Boosting.

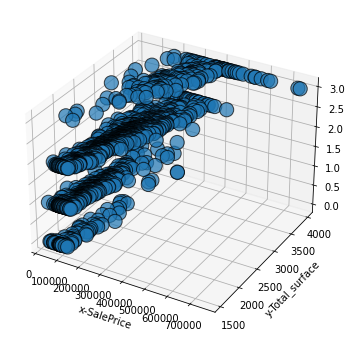
This is the pair plot which gives you idea about how other variables change with respect to other variablesA picture containing text, crossword puzzle

Description automatically generated

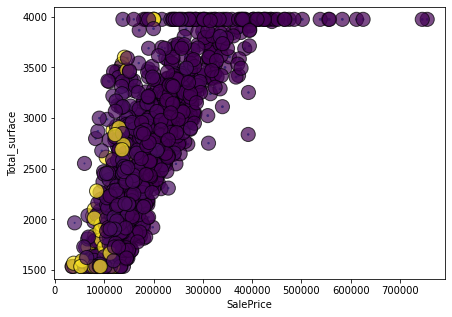
This graph gives the visual representation to identify outliers.Chart, histogram

Description automatically generated

visual correlation between the variables. visual representation of price, area and Garagcars.

Chart

Description automatically generated



**Conclusion:** We used Random Forest, Decision Tree and Linear Regression models to train our data. While evaluating the accuracy for the test data, random forest performed the best amongst the others. This model can be used to predict, for example, which type of house within Iowa is likely to increase and decrease in price based on various scenarios.

**Future Research Directions:**

As a suggestion for future research direction, we suggest that people take into consideration the features that were deemed as most important, this might help them estimate the house price better. The model can be used also with datasets that cover different cities and areas provided that they contain the same features, thus it could help make a generalized model, which doesn’t get affected based on the city/state selected.

**References:**

*Data source,*https://www.kaggle.com/c/home-data-for-ml-course/data

T. G. Dietterich, An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization, Machine learning 40 (2000), no. 2, 139– 157

ANALYSIS AND PREDICTION OF REAL ESTATE PRICES: A CASE OF THE BOSTON HOUSING MARKET Sharmila Muralidharan, Seattle University, muralidh@seattleu.edu Katrina Phiri, Seattle University, phirik@seattleu.edu Sonal K. Sinha, Seattle University, sinhas1@seattleu.edu Ben Kim, Seattle University, bkim@seattleu.edu