DAYANANDA SAGAR UNIVERSITY

KUDLU GATE, BANGALORE – 560068

Bachelor of Technology

In

COMPUTER SCIENCE AND ENGINEERING

Special Topic- II Report

HOUSE PRICE PREDICTION USING MACHINE LEARNING

By

DARSHAN M – ENG20CS0082

DUGGI REDDY VISWANTH– ENG20CS0095

DANIEL THOMAS – ENG20CS0079

VIVEK BELAGALI – ENG20CS0414

Under the supervision of

Dr. Bondu Venkateswarlu

Associate Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,

SCHOOL OF ENGINEERING

(2022 – 2023)

DAYANANDA SAGAR UNIVERSITY

School of Engineering

Department of Computer Science& Engineering

Kudlu Gate, Bangalore –560068
Karnataka, India

CERTIFICATE

This is to certify that the Special Topic II titled "**HOUSE PRICE PREDICTION USING MACHINE LEARNING"** is carried out by **DUGGI VISHWANTH REDDY – ENG20CS0095 , DARSHAN M – ENG20CS0082 , DANIEL THOMAS – ENG20CS0079 VIVEK BELAGALI – ENG20CS0414** bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year **2022-2023**.

Dr. Bondu Venkateswarlu	**Dr. Girisha G S**	**Dr. Udaya Kumar Reddy KR**
Associate Professor	Chairman, CSE	Dean
Dept. of CS&E,	School of Engineering	School of Engineering
School of Engineering	Dayananda Sagar University	Dayananda Sagar University
Dayananda Sagar University		

Name of the Examiner

Signature of Examiner

DECLARATION

We**, DUGGI VISHWANTH REDDY – ENG20CS0095, DARSHAN M – ENG20CS0082, DANIEL THOMAS – ENG20CS0079, VIVEK BELAGALI – ENG20CS0414** are students of the fifth semester B.Tech in **Computer Science and Engineering**, at School of Engineering, **Dayananda Sagar University**, hereby declare that the Special Topic II titled **"HOUSE PRICE PREDICTION USING MACHINE LEARNING"** has been carried out by us and submitted in partial fulfillment for the award of degree in **Bachelor of Technology in Computer Science and Engineering** during the academic year **2022-2023**

Student	**Signature**
DUGGI VISWANTH REDDY **ENG20CS0095**	
DANIEL THOMAS **ENG20CS0079**	
DARSHAN M **ENG20CS0082**	
VIVEK BELAGALI **ENG20CS0414**	

Place: Bangalore
Date :08/12/2022

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this Special Topic II.

First, we take this opportunity to express our sincere gratitude to School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

We would like to thank **Dr. Udaya Kumar Reddy K R, Dean**, **School of Engineering& Technology**, **Dayananda Sagar University** for his constant encouragement and expert advice. It is a matter of immense pleasure to express our sincere thanks to **Dr. Girisha G S , Chairman**, **Department of Computer Science and Engineering**, **Dayananda Sagar University,** for providing the right academic guidance that made our task possible.

We would like to thank our guide **Dr. Bondu Venkateswarlu Associate Professor Dept. of Computer Science and Engineering**, **Dayananda Sagar University**, for her valuable time to extend help in every step of our Special Topic II, which paved the way for smooth progress and the fruitful culmination of the project.

We would like to thank our Special Topic II Coordinators, **Dr. Savitha Hiremath, Dr.Bondu Venkateswarlu** and all the staff members of Computer Science and Engineering for their support.

We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the Special Topic II.

TABLE OF CONTENTS

Page no

ABSTRACT

House Price Index (HPI) is commonly used to estimate the changes in housing price. Since housing price is strongly correlated to other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price. There has been a considerably large number of papers adopting traditional machine learning approaches to predict housing prices accurately, but they rarely concern about the performance of individual models and neglect the less popular yet complex models. As a result, to explore various impacts of features on prediction methods, this paper will apply both traditional and advanced machine learning approaches to investigate the difference among several advanced models. This paper will also comprehensively validate multiple techniques in model implementation on regression and provide an optimistic result for housing price prediction.

CHAPTER 1 INTRODUCTION

Earlier, it's a very popular and common practice to price the property without the proper evaluation of the land, infrastructure etc. We need a proper prediction on the real estate and the houses in housing market we can see a mechanism that runs throughout the properties buying and selling buying a house will be a life time goal for most of the individual but, there are lot of people making mistake in India as most of the people are buying properties from the people, they don't know by seeing the news all around them.

In India, people buy properties which are too expensive but it's not worth it. In the housing market 2016 the house sold in India was about 80 lakhs but the real price according locality and size was about 60 lakhs. In earlier year, there was an economic collapse that give the clue to the impending disaster, this situation is currently happening and the prices of houses are getting higher compared to current economic situation of our country, the Indian government fails to produce the data about the houses so it was very difficult for peoples to buy the properties. Therefore, the people searched on internet for the evidence for house price.

Many methods have been used in the price prediction like advance regression in this I am trying to predict the real estate price for the future using the machine learning techniques with the help of the previous works. I have used the multiple regression and more algorithms with different tools to predict the house price. The purpose of this paper is to establish the proper data preprocessing practices in order to increase the accuracy of machine learning algorithms.

1.1. OVERVIEW OF PROJECT

The sales of the houses are determined on various factors like the location, area, population and some of the information to predict the individual housing price. In addition to these housing prices, the prediction of the housing prices can greatly assist in the prediction of the future housing prices of the real estate. This study uses the machine learning algorithms and technology as a research methodology to develop a housing price prediction. Many algorithms are used here to effectively increase the percentage of the prediction which is considered as the best models in the price prediction. This project shows us that the machine learning algorithm based on accuracy, consistency out performs the other in the performance of the housing price prediction. The project can be created using python (AI/ML), HTML, CSS, JSS. Python is used for writing the Machine Learning Algorithms .HTML, CSS and JS is used for designing the front end of the system. At last, I can conclude y saying that House Price Prediction system will be very helpful in detecting the prices of the houses and keeping the record of the high and low of the prices. So, it will help the user to know the real price of the property, it could not be used for any fraud means.

1.2. SCOPE OF THE PROJECT

In this project, house prices will be predicted given explanatory variables that cover many aspects of residential houses. The goal of this project is to create a regression model that is able to accurately estimate the price of the house given the features. In this dataset made for predicting the Boston House Price Prediction.

We all know that a house price is a number from some defined assortment, so obviously prediction of prices of houses is a regression task. To forecast house prices one person usually tries to locate similar properties in his or her neighborhood and based on collected data that person will try to predict the house price.

CHAPTER 2 PROBLEM DEFINITION

Houses are one of the necessary needs of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world's economy. It is a very large market and there are various companies working in the domain. Data science comes as a very important tool to solve problems in the domain to help the companies increase their overall revenue, profits, improving their marketing strategies and focusing on changing trends in house sales and purchases. Predictive modelling, Market mix modelling, recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies. Our problem is related to one such housing company.

A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia. The data is provided in the CSV file below. The company is looking at prospective properties to buy houses to enter the market.

CHAPTER 3 LITERATURE SURVEY

- [1] P. Durganjali, et al., proposed a house resale price prediction using classification algorithms. In this paper, the resale price prediction of the house is done using different classification algorithms like Logistic regression, Decision tree, Naive Bayes and Random Forest is used and we use International Journal of Computer Sciences and Engineering Vol.8(7), July 2020, E-ISSN: 2347-2693 © 2020, IJCSE All Rights Reserved 99 AdaBoost algorithm for boosting up the weak learners to strong learners. Several factors that are affecting the house resale price includes the physical attributes, location as well as several economic factors persuading at that time. Here we consider accuracy as the performance metrics for different datasets and these algorithms are applied and compared to discover the most appropriate method that can be used the reference for determining the resale price by the sellers.

- [2] Ayush Varma, et al., proposed house price prediction using machine learning and neural networks. Housing prices keep changing day in and day out and sometimes are hyped rather than being based on valuation. Predicting housing prices with real factors is the main crux of our research project. Here we aim to make our evaluations based on every basic parameter that is considered while determining the price. We use various regression techniques in this pathway, and our results are not sole determination of one technique rather it is the weighted mean of various techniques to give most accurate results. The results proved that this approach yields minimum error and maximum accuracy than individual algorithms applied. We also propose to use real-time neighborhood details using Google maps to get exact real-world valuations.

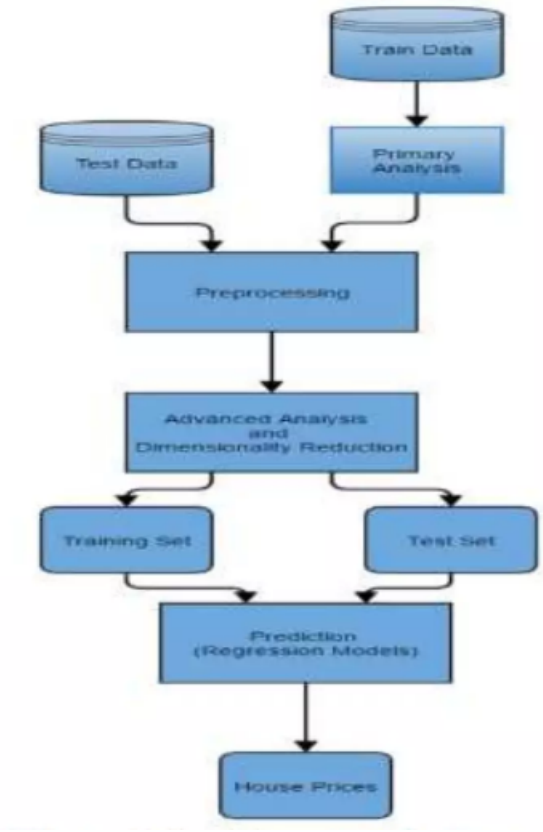
- [3] Sifei Lu, et al., proposed a hybrid regression technique for house prices prediction. With limited dataset and data features, a practical and composite data pre-processing, creative feature engineering method is examined in this paper. The paper also proposes a hybrid Lasso and Gradient boosting regression model to predict individual house price. The proposed approach has recently been deployed as the key kernel for Kaggle Challenge “House Prices: Advanced Regression Techniques”. The performance is promising as our latest score was ranked top 1% out of all competition teams and individuals.

CHAPTER 4 PROJECT DESCRIPTION

In this project we will create a machine learning model with linear regression on housing dataset. This model will aid us in making better real estate decisions by making house price predictions in that area. For this we will carry out data exploration for better understanding of data and will require preprocessing to improve the model accuracy.

In machine learning we write computer programs which automatically improve with experience which are termed as machine learning models. It saves us from explicitly writing code for complex real-world data. In this project we are going to use supervised learning, which is a branch of machine learning where we teach our model by examples. Here we will first explore different attributes of Boston housing dataset then a part of dataset will be used to train the linear regression algorithm after that we will use the trained model to give predictions on remaining part of dataset.

4.1. PROPOSED DESIGN



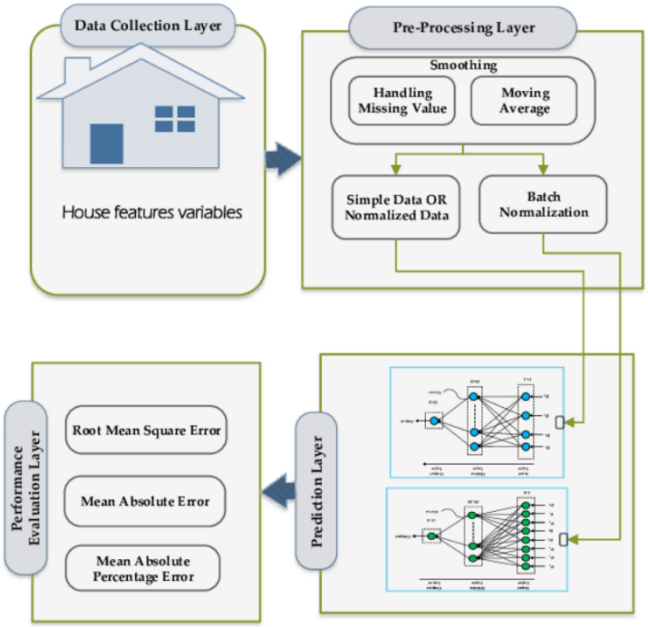
Stage 1: Collection of data Information handling strategies and cycles are various. We gathered the information for Mumbai's land properties from different land sites. The information would have traits, for example, Location, cover region, developed region, age of the property, postal district, and so forth We should gather the quantitative information which is organized and ordered. Information assortment is required before any sort of AI research is completed. Dataset legitimacy is an unquestionable requirement in any case it is a waste of time to break down the information.

Stage 2: Information preprocessing Data preprocessing is the most well-known approach to cleaning our instructive file. There might be missing characteristics or irregularities in the dataset. Data cleansing can help with these issues. Expecting a variable to have a large number of missing attributes we drop persons characteristics else substitute them with the typical worth.

Stage 3: The model's education We should train the model first since the data is separated into two modules: a Training set and a Test set. The objective variable is

joined by the readiness set. The layout of educational assortment is computed using a decision tree regressor. A backslide model is collected as a tree structure by the Decision tree.

Stage 4: Testing and Integrating with UI The test dataset is fed into the pre-programmed model, and home expenses are predicted. The front end, which includes Flask in Python, is then designed using the pre-arranged model.



CHAPTER 5 REQUIREMENTS

5.1 Functional Requirements:

USER INTERFACE: The user interface will be a website. The user has to enter all the attributes correctly and in the required format.

PROPER FORECASTING: The system has to properly predict the price of the house according to the input given by the user.

RECOMMENDATION SYSTEM: According to the input given by the user, the recommendation system will recommend the best property.

DATABASE: Dataset should contain large number of entities so that it will increase the accuracy of the predicted price and suggest a better property.

5.2. NON - FUNCTIONAL REQUIREMENTS

The non-functional requirement elaborates a performance characteristic of the application. Non-functional requirements impose constraints on the design or implementation (such as performance).

The following are the functional requirements for our application as listed below:

- Performance- The response time is short for the requested service.The application provides user friendly interface.
- Reliability- The application is highly realistic and generates all update information in correct order.
- Security- The application is secured and authentication is provided by proper login.
- Availability- The application is available all the time.

5.3. HARDWARE REQUIREMENTS

Hardware requirements is most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware, A hardware requirements list is often accompanied by a hardware compatibility list (HCL), especially in case of operating systems. An HCL lists tested, compatible, and sometimes incompatible hardware devices for a particular operating system or application. The following sub-sections discuss the various aspects of hardware requirements. Hardware that is required for this project can be listed as follows:

- Laptop/PC for Application Development.
- Server (Windows 7/8/10 (32-bit or 64-bit)).
- Local storage

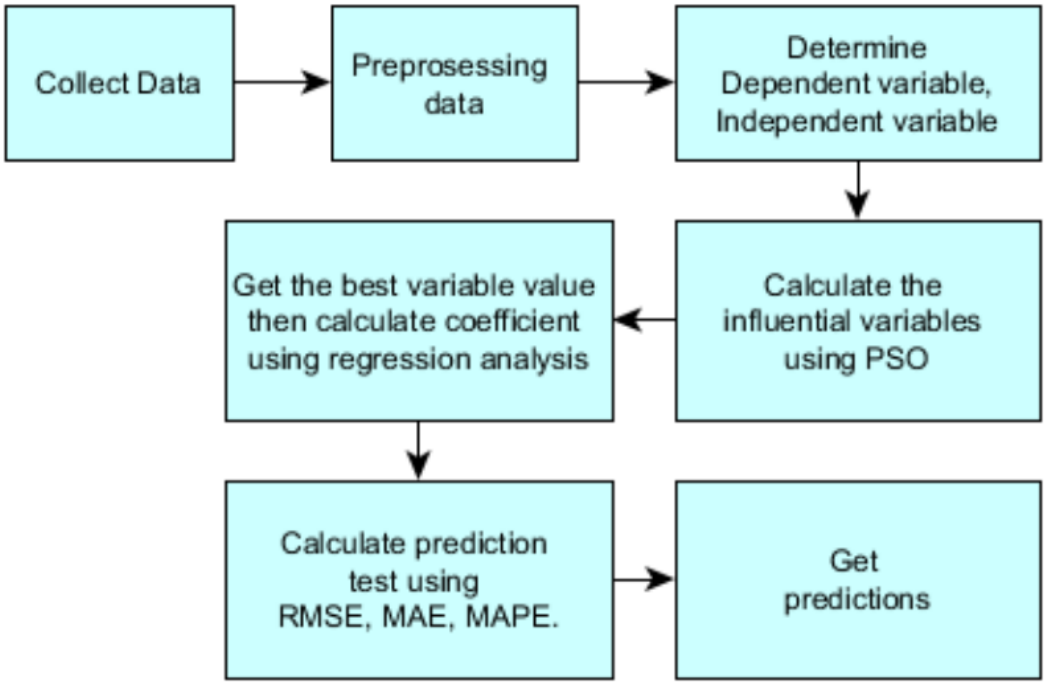
5.4. SOFTWARE REQUIREMENTS

Software requirements deal with defining software resource requirements and prerequisites that need to be installed on a computer to provide optimal functioning of an application. These requirements or prerequisites are generally not included in the software installation package and need to be installed. Software that are required for this project can be listed as follows

- Python.
- Jupyter Notebook

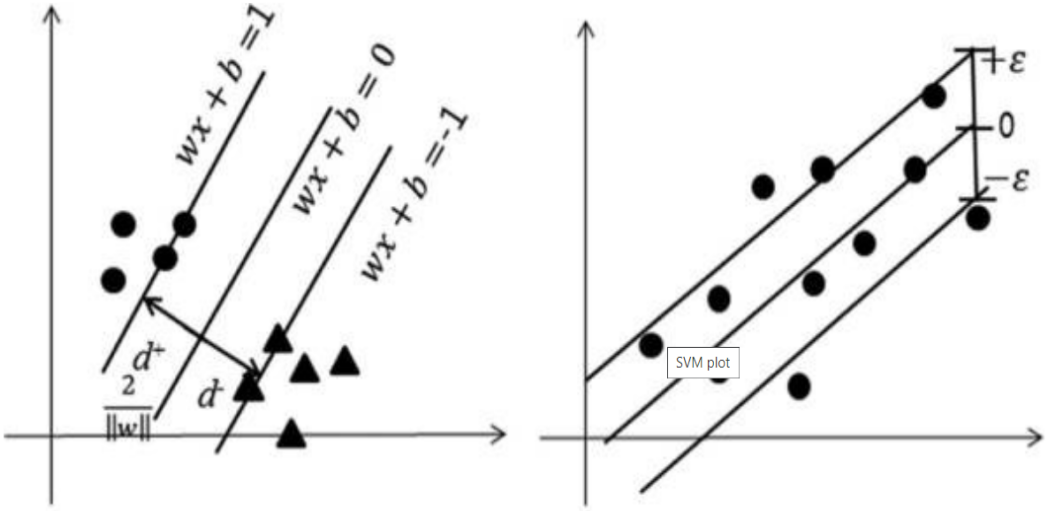
CHAPTER 6 METHODOLGY

- Importing required libraries.
- Reading our input data for House Price Prediction.
- Describing our data.
- Analysing information from our data.
- Plots to visualize data of House Price Prediction.
- Scaling our data.
- Splitting our data for training and test purposes.
- Training our Linear Regression model for House Price Prediction.
- Let’s visualize our predictions of House Price Prediction.
- Plotting the residuals of our House Price Prediction model.
- Observe the coefficients.



CHAPTER 7 EXPERIMENTATION

Support Vector Machine is a supervised machine learning algorithm that can be used for both classification and regression tasks. Support Vector Regression works on the same principle as that of an SVM. The important idea behind the SVR is to find the best-fit hyper-plane along with the optimum threshold. The optimum threshold value is the distance between the hyperplane and the boundary line. The boundary line is defined by the support vectors that lie farthest to the hyperplane. It is similar to fit the best fit line along with some threshold limits. The threshold limit is the distance between the hyperplane and the boundary line.



(a) Support Vector Classification (b) Support Vector Regression

Following are some characteristics of SVR:

- SVR is sensitive to outliers (for Hard Margin only)
- Considers outermost data points for the determination of margin

- SVR is insensitive to inner data points
- SVR can tackle non-linear data using the Kernel Trick

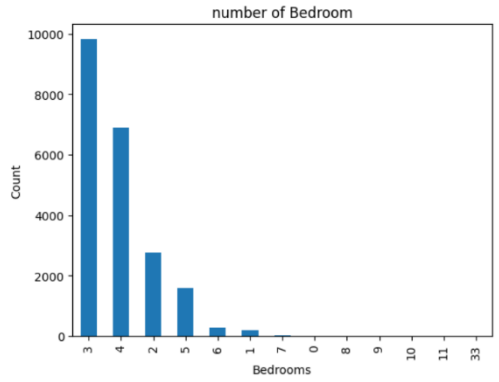
Result: 60% accuracy

CHAPTER 8 TESTING & RESULTS

Case 1 -Number of bedrooms

```
In [8]: data['bedrooms'].value_counts().plot(kind='bar')
        plt.title('number of Bedroom')
        plt.xlabel('Bedrooms')
        plt.ylabel('Count')
        sns.despine

Out[8]: <function seaborn.utils.despine(fig=None, ax=None, top=True, right=True, left=False, bottom=False, offset=None, trim=False)>
```

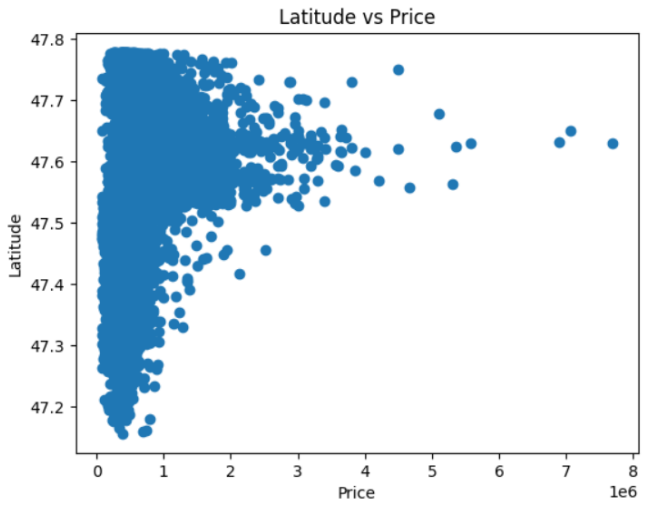


- If you're in the market for a new home, chances are the number of bedrooms the home has is a big factor in your search. Watch Scott Higashi explain how number of bedrooms affects the sales, price of single-family homes and condos. Not surprisingly, median home prices increase for both single-family homes and condos with each additional bedroom.
- The Sold-to-List price ratio is 100 percent for both 2- and 3- bedroom condos and single-family homes, another indicator of strong demand for those types of homes. When we look at studios, 1-bedroom condos and 4- and 5- bedroom homes, the Sold-to-List price ratio decreases slightly, meaning some homes sold for less than the asking price

Case 2: -Latitude vs Price

```
In [12]: plt.scatter(data.price,data.lat)
         plt.xlabel("Price")
         plt.ylabel('Latitude')
         plt.title("Latitude vs Price")
```

```
Out[12]: Text(0.5, 1.0, 'Latitude vs Price')
```

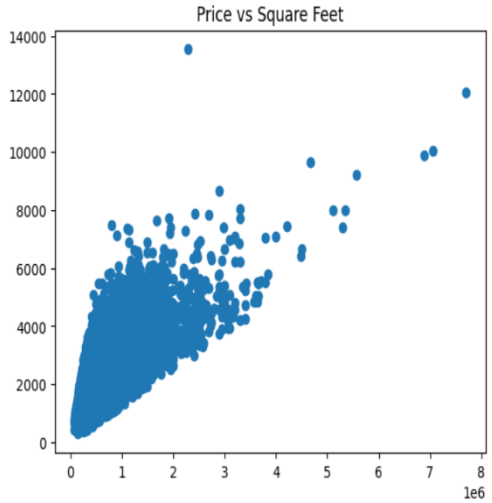


Based on lat/lon, tree-based methods divide the map in rectangular pieces. The stronger the effect and the more data in a certain area, the smaller the pieces. In less dense regions, the pieces would not be too small.

You would not add them as linear effects in a linear regression. There, you would need to consider different approaches. A simple would be to represent lat/lon each by a cubic spline and add interaction terms between them.

Case 3:Price vs Square feet

```
In [10]: plt.scatter(data.price,data.sqft_living)
         plt.title("Price vs Square Feet")
Out[10]: Text(0.5, 1.0, 'Price vs Square Feet')
```



As we can see from all the above representation that many factors are affecting the prices of the house , like square feet which increases the price of the house and even location influencing the prices of the house.

The plot that we used above is called scatter plot , scatter plot helps us to see how our data points are scattered and are usually used for two variables. From the first figure we can see that more the living area , more the price though data is concentrated towards a particular price zone , but from the figure we can see that the data points seem to be in linear direction. Thanks to scatter plot we can also see some irregularities that the house with the highest square feet was sold for very less , maybe there is another factor or probably the data must be wrong. The second figure tells us about the location of the houses in terms of longitude and it gives us quite an interesting observation that -122.2 to -122.4 sells houses at much higher.

CONCLUSION

We have walked through setting up basic simple linear model to predict housing prices resulting from macroeconomic forces and how to assess the quality of a linear regression model on a basic level.

To be sure, explaining housing prices is a difficult problem. There are many more predictor variables that could be used. And causality could run the other way; that is, housing prices could be driving our macroeconomic variables; and even more complex still, these variables could be influencing each other simultaneously.

I encourage you to dig into the data and tweak this model by adding and removing variables while remembering the importance of OLS assumptions and the regression results.

Most importantly, know that the modeling process, being based in science, is as follows: test, analyze, fail, and test some more.

REFERNCES

- House Price Prediction Using Machine Learning. Naga Satish, Ch. V. Raghavendra, M.D. Sugnana Rao, Ch.Srinivasulu|International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019
- House Price Prediction Bindu Sivasankar1, Arun P. Ashok2, Gouri Madhu3, Fousiya S.|International Journal of Computer Sciences and EngineeringVol.8, Issue.7, July 2020 E-ISSN: 2347-2693
- House Price Prediction Using Machine Learning Anand G. Rawool, Dattatray V. Rogye, Sainath G. Rane, Dr. Vinayk A. Bharadi|IRE Journals, Volume 4 Issue 11, MAY 2021