**HOUSE PRICE PREDICTION USING**

**LINEAR REGRESSION**

**A MINOR PROJECT REPORT**

***Submitted by***

**G.V SAI VARUN [RA1811003020102]**

**M. SATHVIK** **[RA1811003020101]**

**K. YASWANTH KUMAR [RA1811003020072]**

*Under the guidance of*

# Ms. Narmatha K

(Designation , Department of Computer Science and Engineering)

*in fulfillment for the award of the degree*

# BACHELOR OF TECHNOLOGY

*in*

**COMPUTER SCIENCE AND ENGINEERING**

*of*

**FACULTY OF ENGINEERING AND TECHNOLOGY**



**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**RAMAPURAM CAMPUS, CHENNAI -600089**

**NOV 2021**

**(Deemed to be University U/S 3 of UGC Act, 1956)**

# BONAFIDE CERTIFICATE

Certified that this project report titled “**HOUSE PRICE PREDICTION USING LINEAR REGRESSION**” is the bonafide work of **G.VSAI VARUN(RA1811003020102), M.SATHVIK(RA1811003020101), K.YASWANTH KUMAR(RA1811003020072)** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an occasion on this or any other candidate.

|  |  |
| --- | --- |
| SIGNATURE  **Ms.NARMATHA K**  **Assistant Professor,**  Computer Science and Engineering,  SRM Institute of Science and Technology, Ramapuram Campus, Chennai. | SIGNATURE  **Dr. K.RAJA,M.E., Ph.D.,**  **Professor and Head**  Computer Science and Engineering,  SRM Institute of Science and Technology, Ramapuram Campus, Chennai. |

Submitted for the project viva-voce held on at SRM Institute of Science

and Technology , Ramapuram Campus, Chennai -600089.

**INTERNAL EXAMINER EXTERNAL EXAMINE**

# RAMAPURAM,CHENNAI- 89 DECLARATION

We hereby declare that the entire work contained in this project report titled “**HOUSE PRICE PREDICTION USING LINEAR REGRESSION**” has been carried out by **G.V SAI VARUN(RA1811003020102, M.SATHVIK(RA1811003020101), K.YASWANTH KUMAR(RA1811003020072**) at SRM Institute of Science and Technology, Ramapuram Campus, Chennai- 600089, under the guidance of **Ms.Narmatha K ., Assistant Professor**, Department of Computer Science and Engineering.

**Place: Chennai**

**G.V.SAI VARUN**

**M.SATHVIK**

**K.YASWANTH KUMAR**

**Date:**

TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Title** | **Page No.** |
| **1** | **Abstract Acknowledgement**  **List of Figures**  **List of Abbreviations**  **INTRODUCTION** | **iv v ix x**  **3-5** |
| 1.1 | Brief Background |  |
| 1.2 | Need for study |  |
| 1.3 | Objectives |  |
| **2** | **LITERATURE SURVEY** | **6-12** |
| 2.1 | Introduction |  |
| 2.2 | Existing System |  |
| 2.3  2.4  2.5 | Issues in Existing System  Proposed System  Advantages of Proposed system |  |
| 2.6 | Literature Survey |  |
| 2.7 | Problem Statement |  |
| 2.8 | Summary of Literature Survey |  |
| **3** | **SPECIFICATIONS** | **13-14** |
| 3.1 | Introduction |  |
| 3.1.1 | Project Scope |  |

* 1. External Interface Requirements
     1. Hardware Interface
     2. Software Interface

1. **CONCEPTS AND DESIGN 15-16**
   1. System Architecture

1. **METHODOLOGY 17-21**
   1. Algorithm

* 1. Algorithm: Refined Substitution

* 1. Modules

1. **SYSTEM IMPLEMENTATION 22-29**
   1. Sample Code
   2. Implementation Screenshots

1. **CONCLUSION AND FUTURE WORK 30-31**

**REFERENCES**

**<ALL the subheadings should be given page No**’**S>**

## ABSTRACT

People looking to buy a new home tend to be more conservative with their budgets and market strategies. The existing system involves calculation of house prices without the necessary prediction about future market trends and price increase. The goal of our project is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted. Real estate is the least transparent industry in our ecosystem. Predicting housing prices with real factors is the main crux of our research project. Everyone wishes for a house that suits their lifestyle and provides amenities according to their needs. House prices increase every year, so there is a need for a system to predict house prices in the future. House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. House prices keep on changing very frequently which proves that house prices are often exaggerated There are many factors that have to be taken into consideration for predicting house prices such as physical conditions, concept, location, number of rooms, carpet area, how old the property is and other basic local amenities. We will be using Linear Regression Model in Machine Learning. Linear Regression is a Supervised Machine Learning Model for finding the relationship between independent variables and dependent variable by using this algorithm we can predict house prices.

Machine Learning is seeing its growth more rapidly in this decade. Many applications and algorithms evolve in Machine Learning day to day. One such application found in journals is house price prediction. House prices are increasing every year which has necessitated the modeling of house price prediction. These models constructed, help the customers to purchase a house suitable for their need. Proposed work makes use of the attributes or features of the houses such as number of bedrooms available in the house, age of the house, travelling facility from the location, school facility available nearby the houses and Shopping malls available nearby the house location.

House price prediction on a data set has been done by using all the above mentioned

techniques to find out the best among them. The motive of this paper is to help the seller to estimate the selling cost of a house perfectly and to help people to predict the exact time slap to accumulate a house. Some of the related factors that impact the cost were also taken into considerations such as physical conditions, concept and location etc.

ACKNOWLEDGEMENT

### **We  place    on    record    our    deep    sense    of    gratitude    to    our    lionized Chairman**

**Dr.R.SHIVAKUMAR** for providing us with the requisite infrastructure throughout the course.

We take the opportunity to extend our hearty and sincere thanks to our Dean, **Dr.M.MURALI KRISHNA, B.E., M.Tech., Ph.D. MISTE,FIE,C.Engg.,** for manoeuvring us into accomplishing the project.

### **We take the privilege to extend our hearty and sincere gratitude to  the  Professor and Head of the Department,** Dr.K.RAJA,M.E.,PhD., **for his suggestions, support and encouragement towards the completion of the project with perfection.**

We express our hearty and sincere thanks to our guide **Ms. Surya With Assistant Professor,** Computer Science and Engineering Department for her/his encouragement, consecutive criticism and constant guidance throughout this project work.

Our thanks to the teaching and non-teaching staff of the Computer Science and Engineering Department of SRM Institute of Science and Technology, Ramapuram Campus, for providing necessary resources for our project.

#### **PROJECT STUDENT NAMES**

G.V.Sai Varun

M.Sathvik

K.Yashwanth Kumar

LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **S.NO** | **FIGURE NAME** | **PAGE** |
| 1. | Linear Regression Graph | 17 |
| 2. | System Architecture Diagram | 15 |
| 3. | Module Split-up | 21 |

LIST OF ABBREVATIONS

1. Machine Learning (ML)
2. Multiple Linear Regression (MLR)
3. Extreme Gradient Boosting (XGBoost)
4. Support vector machines (SVM)
5. Elastic Net Regression (ER)
6. Lasso Regression (LR)
7. Ridge Regression(RR)
8. Root Mean Square(RMS)

### CHAPTER 1

### INTRODUCTION

**1.1 Brief Background**

Data analysis is defined as a process of cleaning, transforming, and modelling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis.

There are several types of Data Analysis techniques that exist based on business and technology However, the major types of data analysis are:

* + - Text Analysis
    - Statistical Analysis
    - Diagnostic Analysis
    - Predictive Analysis
    - Prescriptive Analysis

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Multiple Linear Regression : Multiple Linear Regression (MLR) is a supervised technique used to estimate the relationship between one dependent variable and more than one independent variables. Identifying the correlation and its cause-effect helps to make predictions by using these relations.

Regularised regression plays a significant part in Multiple Linear Regression because it helps to reduce variance at the cost of introducing some bias, avoid the overfitting problem and solve ordinary least squares (OLS) problems. There are two types of regularisation techniques L1 norm (least absolute deviations) and L2 norm (least squares). L1 and L2 have different cost functions regarding model complexity. Multiple Linear Regression is an extension of Simple Linear Regression and assume that there is a linear relationship between a dependent variable Y and independent variables X.

## Machine learning algorithms are often categorized as supervised or unsupervised:

**Supervised machine learning** Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

**Unsupervised machine learning** These are used when the information used to train is neither classified nor labelled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data. The system explores the data and can draw inferences from datasets to describe hidden structures from unlabelled data.

**Semi-supervised machine learning** These algorithms fall somewhere in between supervised and unsupervised learning, since they use both labelled and unlabelled data for training – typically a small amount of labelled data 23 and a large amount of unlabelled data. Usually, semi-supervised learning is chosen when the acquired labelled data requires skilled and relevant resources in order to train it / learn from it.

**1.2 NEED FOR STUDY:**

Prediction house prices are expected to help people who plan to buy a house so they can know the price range in the future, then they can plan their finance well then they can plan their finance well. In addition, house price predictions are also beneficial or property investors to know the trend of housing prices in a certain location. House prices increase every year, so there is a need for a system to predict house prices in the future. House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. There are several factors that affect house prices. They are dividing these factors into three main groups, there are physical condition, concept and location. Physical conditions are properties possessed by a house, including the size of the house, the number of bedrooms, the availability of kitchen and garage, the availability of the garden, the area of land and buildings, and the age of the house , while the concept is an idea offered by developers who can attract potential buyers, for example, the concept of a minimalist home, healthy and green environment, and elite environment. Location is an important factor in shaping the price of a house. This is because the location determines the prevailing land price. In addition, the location also determines the ease of access to public facilities, such as schools, campus, hospitals and health centres, as well as family recreation facilities such as malls, culinary tours, or even offer a beautiful scenery.

**1.3 OBJECTIVES:**

1. Predict the house price.

1. Using two different models in terms of minimizing the difference between predicted and actual rating
2. The goal of this statistical analysis is to help us understand the relationship between house features and how these variables are used to predict house price.
3. In this project House Price Prediction is implemented using linear regression. Linear regression helps to find the relation between dependent and independent variable this helps to predict the accurate result.

**CHAPTER 2**

**LITERATURE REVIEW:**

**2.1 Introduction:**

We provide a brief overview of the existing reviews that have studied House Price Prediction.

While several literature reviews/surveys are published in order to study the House Price Prediction, there are still some ongoing challenges that have not been addressed. Table 1 presents a comparative summary of the existing House Price Prediction reviews/surveys according to six criteria, namely proposing a taxonomy, considering several prices dataset, considering application domains, covering Price Prediction improvement tools, identifying research gaps, and scope of literature review. We observe that there is a lack of taxonomy focusing on Price Prediction improvement (i.e., addressing Price Prediction, security, privacy, and performance issues) and Price Prediction usage (i.e., addressing domain-specific issues).

* 1. **Existing System:**

1. In today’s world, everyone wishes for a house that suits their lifestyle and provides amenities according to their needs.
2. The present method is that the customer approaches a real estate agent to manage his/her investments and suggest suitable estates for his investments.
3. There are many factors that have to be taken into consideration for predicting house prices such as location, number of rooms, carpet area, how old the property is and other basic local amenities.

**2.3 Issues in Existing System:**

1.It is not possible to predict the future prices of homes mentioned by the customer. Because of this, investing in an apartment increases the risk or part of it significantly.2. Expensive property leads to an increase in rents – because rented housing is a substitute for owner-occupation. 3. Rapid house price inflation lead to unsustainable asset which leads to macroeconomic instability.4. The real estate agent method is risky as the agent might predict wrong estates and thus leading to loss of the customers’ investments.

* 1. **Proposed System:**

The objective of this project is to predict the house prices so as to minimize the problems faced by the customer. The algorithm for house price prediction are Data analysis and Linear regression.The goal of this statistical analysis is to help us understand the relationship between house features and how these variables are used to predict house price.To accomplish this task, the python programming language is used. Python is a high-level programming language for general purpose programming. Saves time and energy.Easy to access the system anywhere and anytime.House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house.The linear regression algorithm helps to fulfil customers by increasing the accuracy of estate choice and reducing the risk of investing in an estate.

**2.4 Literature Survey**

## REVIEW 1:

**TITLE:** House Price Prediction Using Machine Learning And Neural networks

**JOURNAL:** INSPEC

**CONCEPT:** Predicting the house price using of neural network along with Boosted Regression was used.

**DISADVANTAGE:**The accuracy of the system can be improved. Several more cites can be included in the system if the size and computational power increases of the system.

## REVIEW 2:

**TITLE:** Predicting the house price using the simple stacking algorithm

**JOURNAL:** IEEE

**CONCEPT:** Predicting the house price using the simple stacking algorithm **DISADVANTAGE:** Different algorithms predicted different outputs with different accuracies. REVIEW 3:

**TITLE:** Residential Asset Pricing Prediction using Machine Learning**. JOURNAL:** IEEE

**CONCEPT:** This paper predicts residential housing prices using three machine learning methods, namely, support vector machine, random forest, and regression

**DISADVANTAGE:** However, this paper still has some shortcomings, for example they did not consider housing heterogeneity.

## REVIEW 4:

**TITLE:** House Price Prediction Using Regression Techniques.

**JOURNAL:** IEEE

**CONCEPT:** The paper involves predictions using different Regression techniques like Multiple linear, Ridge, LASSO, Elastic Net, Gradient boosting and Ada Boost Regression.

**DISADVANTAGE:** Data imbalance leads to a decrease in classification accuracy

## REVIEW 5:

**TITLE:** Housing price prediction via improved machine learning

**JOURNAL:** IEEE

**CONCEPT:** This papers involves Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Hybrid Regression and Stacked Generalization.

**DISADVANTAGE:** Stacked Generalization Regression did not work well on the training set as Hybrid Regression.

## REVIEW 6:

**TITLE:** Prediction of House Pricing using Machine Learning with Python

**JOURNAL:** IEEE

**CONCEPT:** Predicting the house price using the simple stacking algorithm

**DISADVANTAGE:**Different alogorithms predicted different outputd with different accuracies.

## REVIEW 7:

**TITLE**: House Price prediction Using Machine Learning and RPA

**JOURNAL:** IRJET

**CONCEPT:** RPA provides to the market have made it one of the top contenders in the current market.It generates more accurate and consistent processes.

## REVIEW 8:

**TITLE**:House **DISADVANTAGE:** Budgetary restrictions are among the biggest reasons.

Price prediction Using Hybrid Regression.

**JOURNAL:** IRJET

**CONCEPT:** Hybrid Regressions result is better than one specific regressions algorithm.There is a need to verify different hybrid combination to get the best score.

**DISADVANTAGE:** Lack of memory ,Local minimization drawback,Complicated architecture

## REVIEW 9:

**TITLE**:House Price prediction based on some economic factors.

**JOURNAL:** IRJETS

**CONCEPT:** The main objective of developing this system is to save time and cost and describe all the possiblity values of house.It does not need much resources.

**DISADVANTAGE:** Places like complexes and multi story apartments are not included.

## REVIEW 10:

**TITLE**:House Price prediction using Data Mining

**JOURNAL:** IJARCCE

**CONCEPT:** In todays real estate world, it has become tough to store such huge data and extra them for ones own requirement. The system makes optimal use of the Data mining Algorithm. The system make use of such data in the most efficient way.

**DISADVANTAGE:** Places like complexes and multi story apartments are not included.

**2.5 Problem statement:**

This is an interesting problem because most of the people will eventually buy/sell a home. This problem allows us, as house price analysts, to learn more about the housing market and helps with making more informed decisions.

**2.6 Summary of Literature Survey:**

To sum up, it can be said that the existing surveys concerning House Price Prediction focus on classifying the documents based on Price Prediction issues. Our work extends the existing surveys by studying the Price Prediction application domains, analyzing the Price Prediction challenges, and introducing some research gaps that need to be addressed in future studies.

The decentralization, auto-enforcing ability, and verifiability characteristics of Price Predictions enable their encoded business rules to be executed in a peer-to-peer network, where each node is “equal”, and none has any special authority without the involvement of a trusted authority or a central server. Thus, Price Predictions are expected to revolutionize many traditional industries, such as financial, healthcare, energy, etc. In this document, we presented a comprehensive survey of House Price Prediction from both technical and usage points of view. Thus, we introduced a taxonomy of existing House Price Prediction solutions, categorized the included research documents, and discussed the existing Price Prediction- based

**CHAPTER 3**

**SPECIFICATIONS**

**3.1 Introduction:**

This article refers together with latest Forecast on Research predictions considering trends to further plan their economics. The main motivation of the project FORECAST VARIATIONS ON HOUSE PRICE was to make the best possible prediction of house prices by using appropriate algorithms and finding out which among them is best suitable for predicting the price with low error rate. This is an interesting problem because most of the people will eventually buy/sell a home. This problem allows us, as house price analysts, to learn more about the housing market and helps with making more informed decisions. The analysis that were done in this document is mainly based on the datasets.

**3.1.1 Project Scope:**

1.Adding estate database of more cities which will provide the user to explore more estates and reach an accurate decision.

2.In-depth details of every property will be added to provide ample details of a desired estate.

3.More factors like recession that affect the house prices shall be added.

4. In this project House Price Prediction is implemented using Linear regression. Linear regression helps to find the relation between dependent and independent variable this helps to predict the accurate result.

**3.2 External Interface Requirements:**

**3.2.1 Hardware Interface:**

Ram : 4 GB Ram or above

Processor : I3 Intel Processor or above

Hard Disk : 6 GB and more

Speed : 1GHZand more

Memory: 100 MB

**3.2.2 Software Interface:**

IDE : Google Colab / Anaconda Navigator -Jupyter Notebook

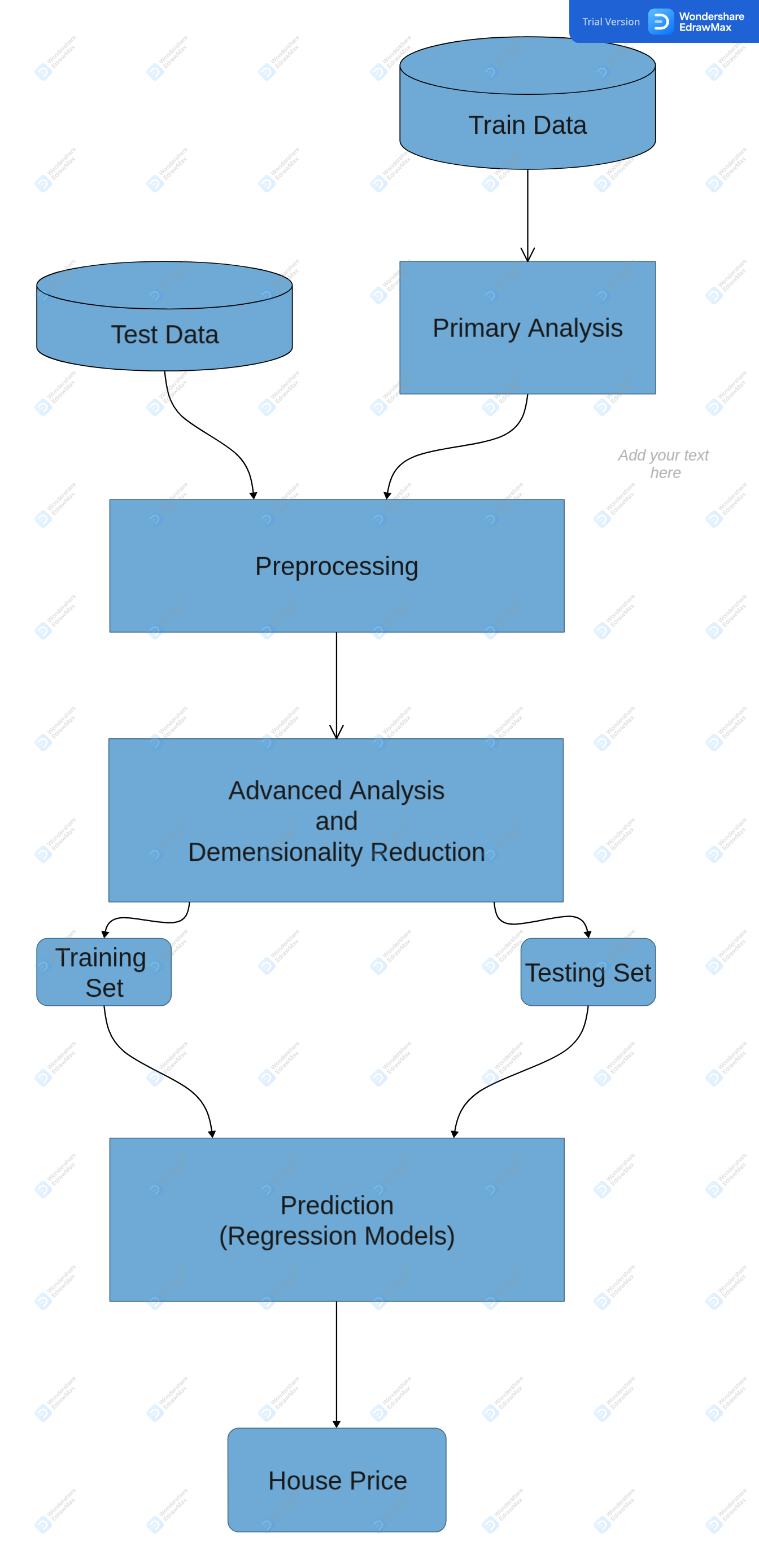
Operating System: Windows 10/8/7

Language: Python

**CHAPTER 4**

**CONCEPTS AND DESIGN**

## SYSTEM ARCHITECTURE

****

We first allow the test data and train data should be pre-processed . Basically, test data is used to validate the build-up model. It is used as unbiased evaluation of a final model to fit in the training data. Test data is data which has been specifically identified for use in tests, typically of a computer program. Some data may be used in a confirmatory way, typically to verify that a given set of input to a given function produces some expected result. Test data is the Input feed for Testing the Application. Test Data helps the developers to find the problem during fixes.

Test Data may be used in a confirmatory way, typically to verify that a given set of input to a given function produces some expected result. While , Train data is used to implement the build model. The training data to The training data is then undergo primary analysis. Analysing primary data is the process of making sense of the collected data to answer research questions or support or reject research hypotheses that a study is originally designed to assess. That is ,it will filter out those data which does not comes under the applied condition.

Once it is done both the test and training is pre-processed .Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. ... And while doing any operation with data, it is mandatory to clean it and put in a formatted way. In this step it filters out both test data and training data.

The pre-processed data will be send to Advanced analysis and Dimensionality reduction. Dimensionality reduction refers to techniques for reducing the number of input variables in training data. Fewer input dimensions often mean correspondingly fewer parameters or a simpler structure in the machine learning model, referred to as degrees of freedom.

Now, the train and test date will get separate and Linear regression model will be applied to the data sets in order to get the accurate output.

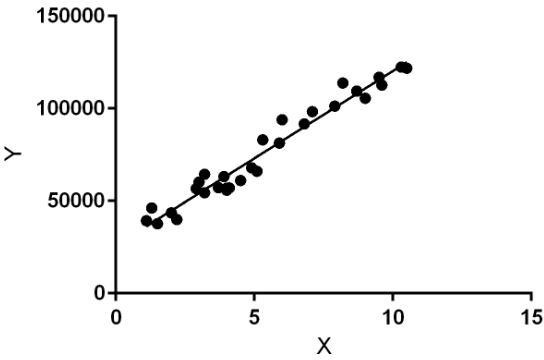
fit the model and testing data to test it

**CHAPTER 5**

**METHODLOGY**

**5.1 Algorithm:**

Linear Regression is a Supervised Machine Learning Model for finding the relationship between independent variables and dependent variable. Linear regression performs the task to predict the response (dependent) variable value (y) based on a given (independent) explanatory variable (x). So, this regression technique finds out a linear relationship between x (input) and y (output)..Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.



Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.

In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.

Hypothesis function for Linear Regression :

image3.jpeg

While training the model we are given :

x: input training data (univariate – one input variable(parameter)) y: labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best θ1 and θ2 values.

The variable we are predicting is called the criterion variable and is referred to as Y. The variable we are basing our predictions on is called the predictor variable and is referred to as X. When there is only one predictor variable, the prediction method is called **Simple Regression.**

## **5.2 Algorithm:Refined Substitution:**

We use train data and test data , train data to train our machine and test data to see if it has learnt the data well or not.

Steps to follow:

1. We import our dependencies , for linear regression we use sklearn (built in python library) and import linear regression from it.
2. We then initialize Linear Regression to a variable reg.Now we know that prices are to be predicted , hence we set labels (output) as price columns and we also convert dates to 1’s and 0’s so that it doesn’t influence our data much . We use 0 for houses which are new that is built after 2014.
3. We again import another dependency to split our data into train and test.
4. I’ve made my train data as 90% and 10% of the data to be my test data , and randomized the splitting of data by using random\_state.
5. So now , we have train data , test data and labels for both let us fit our train and test data into linear regression model.
6. After fitting our data to the model we can check the score of our data .

## **5.3 Modules:**

Data Set Module:

Collection of data is the process of measuring and gathering the information with the help of a software. There are many techniques and procedures to collect the data. We will be collecting the quantitative data which is structured and categorized.

It cleans the data(removal of garbage value) . Data cleansing is the process of cleaning our data set. There could be various garbage values present in the dataset. These garbage values can be removed by checking whether any missing values are present in the data or not. We also need to end the validity of our dataset. Also the values need to be present in a given range. If a variable has many missing values we can drop those values.

Feature Selection Module:

The selection of feature is also known as variable selection. It is a process in which subset selection of parameters or variables from a large universal set of parameters is done. Feature selection techniques are used for three reasons:

* To simplify the model
* To shorten the training time of the model
* To reduce the dimensionality.

Validation of Model:

Validation is the process of checking whether the applied algorithm is the given dataset or not. Thus the accuracy of the model should be as high as possible. After applying the algorithm we can check how well our model is the data. We can also apply two or more models to check the model or which is our dataset the best. The model is viewed as an input-output transformation for these tests. The validation test compares the outputs from the system that is under consideration to the outputs that are obtained from the model provided that same input parameters are given to the model. The output values obtained from the model are recorded.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | image7.png |

download

**CHAPTER 6**

**SYSTEM IMPLEMENTATION**

**6.1 Sample CODE:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

df = pd.read\_csv('C:\\Users\\Tupili Preethi\\OneDrive\Desktop\\USA\_Housing.csv.csv')

df.head()

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

df.info()

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

df.describe()

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

df.columns

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

sns.pairplot(df)

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

sns.distplot(df['Price'])

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

sns.heatmap(df.corr(), annot=True)

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

X = df[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms', 'Avg. Area Number of Bedrooms', 'Area Population']]

y = df['Price']

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.4, random\_state=101)

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

from sklearn.linear\_model import LinearRegression

lm = LinearRegression()

lm.fit(X\_train,y\_train)

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

print(lm.intercept\_)

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

lm.coef\_

#examining the co-efficients of the fitted model.

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

X\_train.columns

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

cdf = pd.DataFrame(data=lm.coef\_.reshape(5,1),index=X\_train.columns,columns=['Coeff'])

cdf

\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

predictions = lm.predict(X\_test)

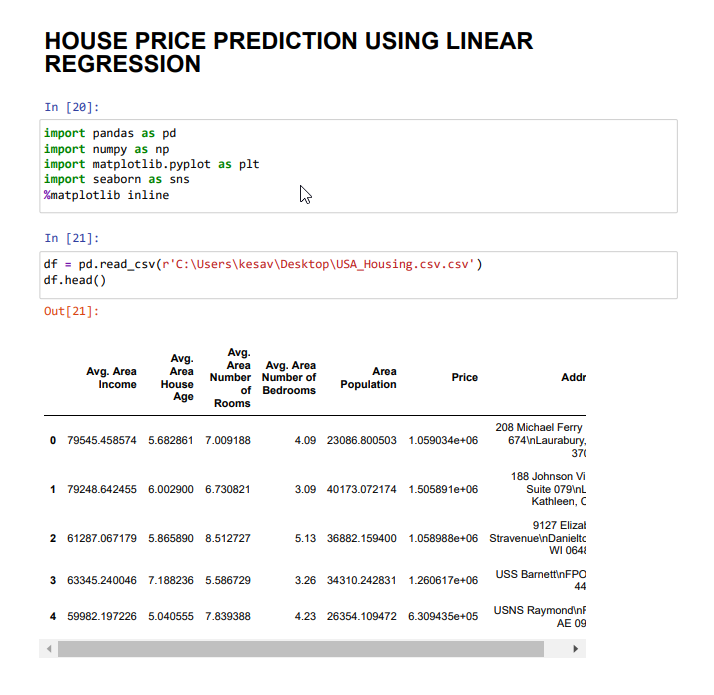
\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

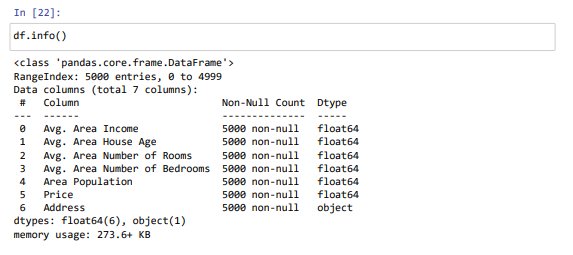
\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

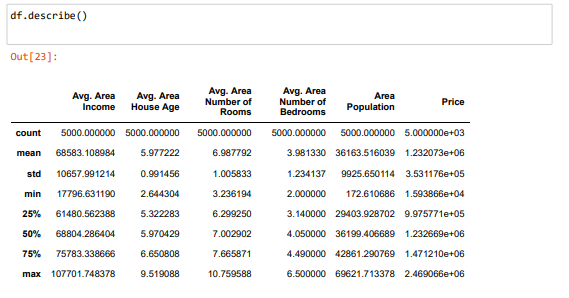
sns.distplot((y\_test-predictions),bins=50);

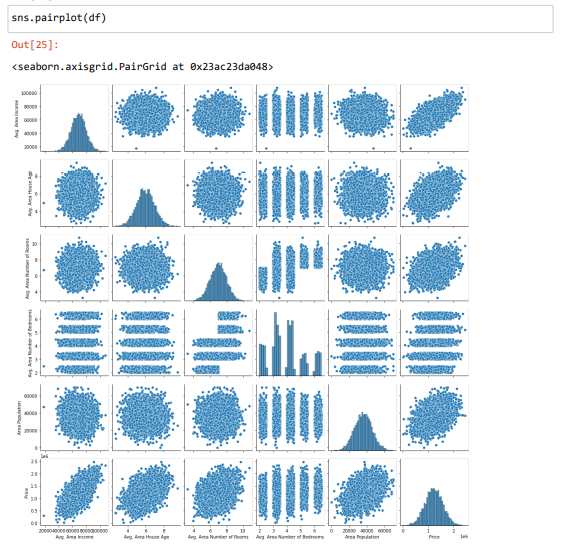
\*\*\*\*\*\*RUN\*\*\*\*\*\*\*\*\*

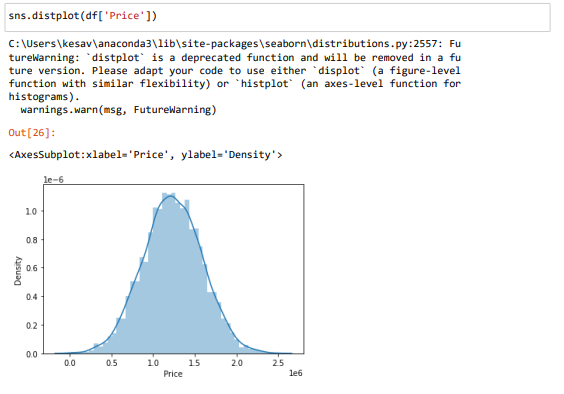
**6.2 Implementation Screenshots:**

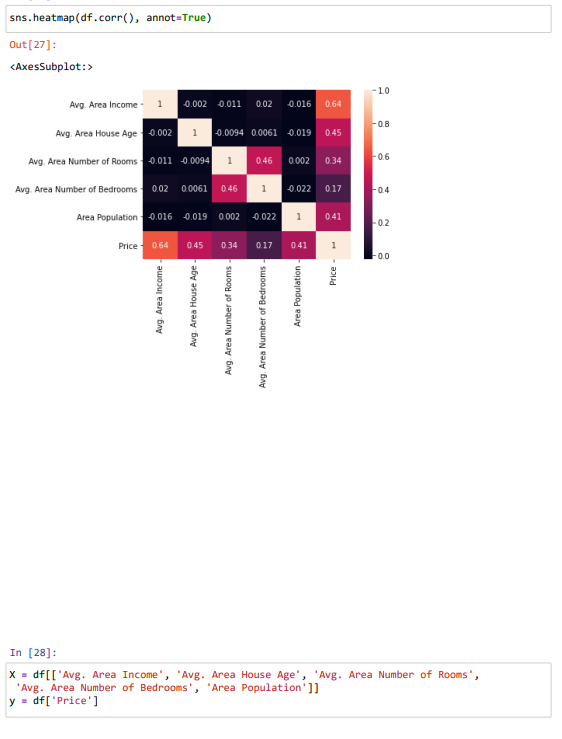
****

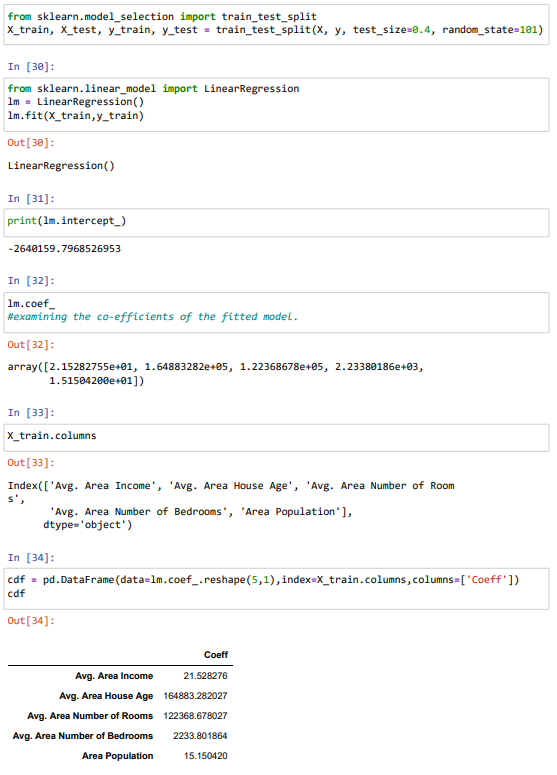
****

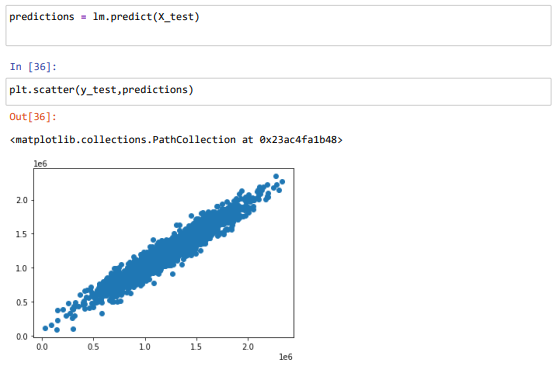


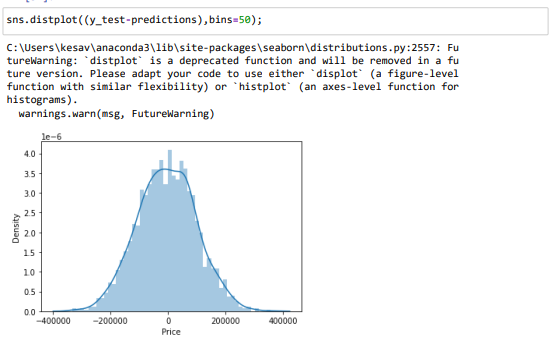












**CHAPTER 7**

**7.1 CONCLUSION:**

In today’s real estate world, it has become tough to store such huge data and extract them for one’s own requirement. Also, the extracted data should be useful. The system makes optimal use of the Linear Regression Algorithm. The system makes use of such data in the most efficient way. The linear regression algorithm helps to fulfil customers by increasing the accuracy of estate choice and reducing the risk of investing in an estate. A lot’s of features that could be added to make the system more widely acceptable. One of the major future scopes is adding estate database of more cities which will provide the user to explore more estates and reach an accurate decision. More factors like recession that affect the house prices shall be added. In-depth details of every property will be added to provide ample details of a desired estate. This will help the system to run on a larger level.

A system that aims to provide an accurate prediction of housing prices has been developed. We have created a Linear Regression Model which we help the real estate agent for estimating the house price.

Although the linear regression model that was created in this post is not perfect, it is able to account for approximately 87.8% of the variation in Sale Price of a property, and it also is able to predict the Sale Price within $23,625.19. Further exploration of the residuals of this model revealed a specific weakness in this model at predicting extreme values. Specifically, the ideal range for predictions appeared to be within $90,000 to $225,000. In order to better predict prices outside of this range, some changes could be made to the model. Ideas for improvement include log transforming features and coding some categorical variables as ordinal.

**REFERENCES:**

1. Ayush Varma , Abhijit Sarma ,Sagar Doshi ,Rohini Nair “House Price Prediction Using Machine Learning and Neural Networks” IEEE 2018.
2. .Mansi Jain , Himani Rajput , Neha Garg , Pronika Chawla “Prediction of House Pricing Using Machine Learning with Python”IEEE 2020, International Conference on Electronics and Sustainable Communication Systems (ICESC)
3. Yiyang Luo “vi” IEEE 2019.
4. CH. Raga Madhuri,G. Anuradha,M. Vani Pujitha”House Price Prediction Using Regression Techniques: A Comparative Study”2019 International Conference on Smart Structures and Systems (ICSSS)
5. Quang Truong, Minh Nguyen, Hy Dang, Bo Mai”Housing Price Prediction via Improved Machine Learning Techniques”2019 International Conference on Identification, Information and Knowledge in the Internet of
6. Things (IIKI2019)
7. Visit “Predict Boston House Prices Using Python & Linear Regression”2019.
8. Vishal Raman, May 2014. Identifying Customer Interest in Real Estate Using Data Mining.
9. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, 2015. Introduction to Linear Regression Analysis
10. Gongzhu Hu, Jinping Wang, and Wenying FengMultivariate Regression Modeling for Home Value Estimates with Evaluation using Maximum Information Coefficient