



**A Mini Project Report**  
**on**  
**“BENGALURU HOUSE PRICE PREDICTION”**

**Submitted in partial fulfilment for the award of the degree of**

**BACHELOR OF TECHNOLOGY (HONOURS)**

**IN**

**COMPUTER SCIENCE (DATA SCIENCE)**

**Submitted by**

**A NAGA KUSHAL**

**19BTRCR047**

**G SAI AVINASH**

**19BTRCR048**

**G NISHEN**

**19BTRCR061**

**M SHAISHTHA**

**19BTRCR014**

**Under the guidance of**

**DR VIAANY**

**Faculty of Engineering & Technology**

**Jain (Deemed-To-Be University)**

**B. Tech (Honours) in Computer Science (Data Science)**

Jain Global Campus, Kanakapura Taluk - 562112

Ramanagara District, Karnataka, India

2021-2022.

## **B. Tech (Honours) in Computer Science (Data Science)**

Jain Global Campus, Kanakapura Taluk - 562112

Ramanagara District, Karnataka, India

### **CERTIFICATE**

This is to certify that the project work titled “**BENGALURU HOUSE PRICE PREDICTION**” is carried out by **A NAGA KUSHAL (19BTRCR047)**, **G SAI AVINASH (19BTRCR048)**, **G NISHEN (19BTRCR061)**, **M SHAISTHA (19BTRCR014)**, a bonafide students of Bachelor of Technology at the Faculty of Engineering & Technology, Jain (Deemed-to-be University), Bangalore in partial fulfilment for the award of degree, Bachelor of Technology (Honours) in Computer Science (Data Science), during the Academic year **2020-2021**.

**DR VIAANY**

Faculty of Engineering &  
Technology,  
Jain (Deemed-to-be University)  
Date:  
Signature:

**Prof. Mohammed Zabeeulla A N**  
**Assistant Professor and**  
**Programme Coordinator,**  
**Dept. of CSE**  
Faculty of Engineering &  
Technology,  
Jain (Deemed-to-be University)  
Date:  
Signature:

**Dr. Devaraj Verma,**  
**Professor and Dy. HoD,**  
**Dept. of CSE**  
Faculty of Engineering &  
Technology,  
Jain (Deemed-to-be  
University)  
Date:  
Signature:

Name of the Examiner    Signature of Examiner

1.

2.

# DECLARATION

We, **A NAGA KUSHAL (19BTRCR047), G SAI AVINASH (19BTRCR048), G NISHEN (19BTRCR061), M SHAISTHA (19BTRCR014),** are students of sixth semester B. Tech (Honours) in **Computer Science (Data Science)**, at Faculty of Engineering & Technology, **Jain (Deemed-To-Be University)**, hereby declare that the project work titled **“BENGALURU HOUSE PRICE PREDICTION ”** has been carried out by us and submitted in partial fulfilment for the award of degree in **Bachelor of Technology (Honours) in Computer Science (Data Science)** during the academic year **2020-2021**. Further, the matter presented in the project has not been submitted previously by anybody for the award of any degree or any diploma to any other University, to the best of our knowledge and faith.

Student Name : A NAGA KUSHAL

Signature

USN:19BTRCR047

Student Name : G SAI AVINASH

Signature

USN: 19BTRCR048

Student Name : G NISHEN

Signature

USN: 19BTRCR061

Student Name : M SHAISHTHA

Signature

USN: 19BTRCR016

Place: Bengaluru

Date:22-06-2022

# ACKNOWLEDGEMENT

*It is a great pleasure for us to acknowledge the assistance and support of a large number of individuals who have been responsible for the successful completion of this project work.*

*First, we take this opportunity to express our sincere gratitude to **Faculty of Engineering & Technology, Jain (Deemed-to-be University)**, for providing us with a great opportunity to pursue our Bachelor's Degree (Honours) in this institution.*

*In particular we would like to thank **Dr. Hariprasad S A, Director, Faculty of Engineering & Technology, Jain (Deemed-to-be University)**, for his constant encouragement and expert advice.*

*It is a matter of immense pleasure to express our sincere thanks to **Dr. Devaraj Verma, Professor and Deputy Head, Department of Computer Science & Engineering, Jain (Deemed-to-be University)**, for providing right academic guidance that made our task possible.*

*It is a matter of immense pleasure to express our sincere thanks to **Prof. Mohammed Zabeeulla, Program Coordinator of Data Science, Dept. of Computer Science & Engineering, Jain (Deemed-to-be University)**, for providing right academic guidance that made our task possible.*

*We would like to thank our guide **DR VIAANY, Dept. of Computer Science & Engineering, Jain (Deemed-to-be University)**, for sparing his valuable time to extend help in every step of our project work, which paved the way for smooth progress and fruitful culmination of the project.*

*We would like to thank our Project Coordinator **Dr. S. Vijayakumar** and all the staff members of Computer Science & 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 completing the Project work successfully.*

*Signature of Students*

# TABLE OF CONTENTS

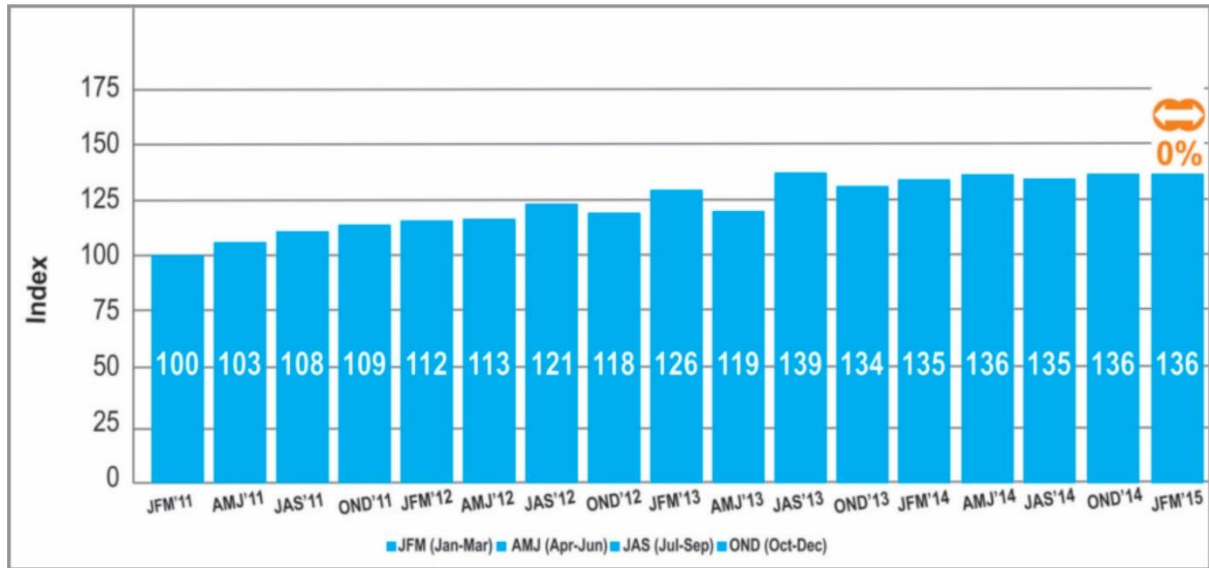
	Page No
<b>CERTIFICATE</b>	ii
<b>DECLARATION</b>	iii
<b>ACKNOWLEDGEMENT</b>	iv
<b>TABLE OF CONTENT</b>	v
<b>ABSTRACT</b>	vi
<b>LIST OF FIGURES</b>	vii
<b>LIST OF TABLES</b>	viii
<b>NOMENCLATURE USED</b>	viii
<b>Chapter 1</b>	1
<b>1. INTRODUCTION</b>	1
1.1. Overview	1
1.2. Problem Definition	1
1.3. Objectives	1
1.4. Methodology	1
1.5. Hardware and Software Tools Used	1
<b>Chapter 2</b>	2
<b>2. LITERATURE SURVEY</b>	2
2.1. Related Work	2
2.2. Existing System	2
2.3. Limitation of Existing System	2
2.4. Proposed System	2
<b>Chapter 3</b>	3
<b>3. METHODOLOGY</b>	3

3.1. Dataset	3
3.2. Architecture	3
3.3. Sequence Diagram	3
<b>Chapter 4</b>	4
<b>4. TOOL DESCRIPTION</b>	4
4.1. Hardware Requirements	4
4.2. Software Requirements	4
<b>Chapter 5</b>	5
<b>5. IMPLEMENTATION</b>	5
<b>Chapter 6</b>	6
<b>6. RESULTS AND ANALYSIS</b>	6
6.1. Result Discussion	6
6.2. Comparison with Previous Studies	6
6.3. Analysis	6
<b>Chapter 7</b>	7
<b>7. CONCLUSIONS AND FUTURE SCOPE</b>	7
<b>REFERENCES</b>	8
<b>APPENDIX</b>	9

## **ABSTRACT**

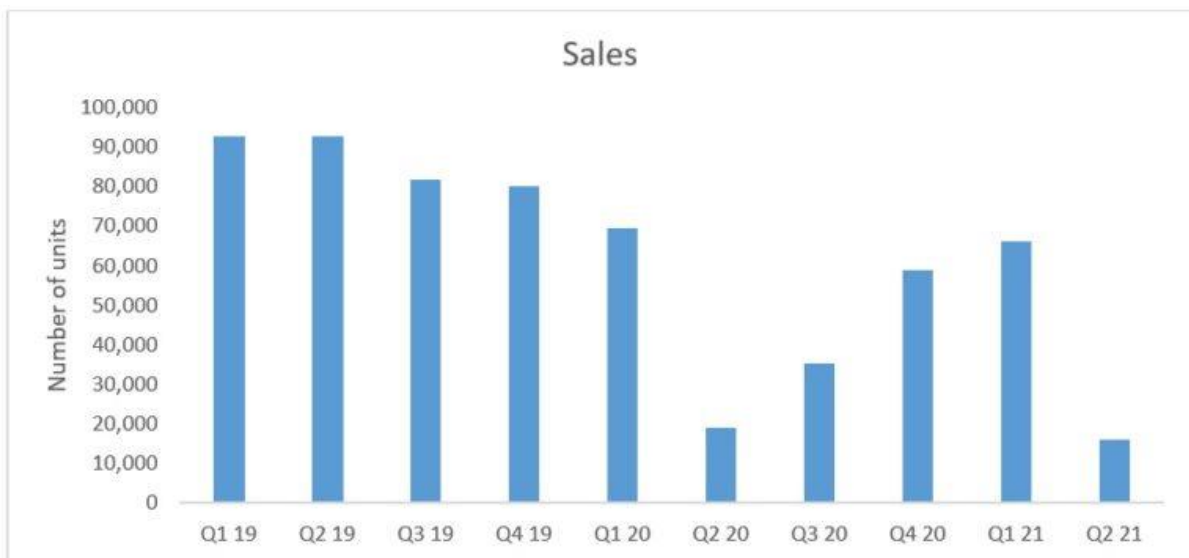
In recent years, Karnataka, one of the hotspots for real estate development, has seen an increase in demand from potential home buyers and investors, and is expected to witness a further boom in the sector by 2020. What are the things that a potential home buyer considers before purchasing a house? The location, the size of the property, vicinity to offices, schools, parks, restaurants, hospitals or the stereotypical white picket fence? What about the most important factor — the price? Now with the lingering impact of COVID-19, the enforcement of the Real Estate (Regulation and Development) Act (RERA), and the lack of trust in property developers in the city, housing units sold across India in 2019 dropped by 3 per cent. In fact, the property prices in Bengaluru fell by almost 5 per cent in the second half of 2019, said a study published by property consultancy Knight Frank. Buying a home, especially in a city like Bengaluru, is a tricky choice. While the major factors are usually the same for all metros, there are others to be considered for the Silicon Valley of India. With its help millennial crowd, vibrant culture, great climate and a slew of job opportunities, it is difficult to ascertain the price of a house in Bengaluru. This paper reflects the effort towards solving the problems mentioned. In this paper, the authors have tried to create such a system which will in turn, give a very accurate prediction about the prices of the house in the city of Bengaluru. The authors have tried to create a user friendly interface design which will enable the users to choose their options as per their requirements and get the estimated price of the house according to their needs.

# PROPINDEX - BENGALURU



**magicbricks**.com

FIG : 1.1



Source: DataLabs, PropTiger Research

FIG : 1.2





## LIST OF FIGURES

<b>Fig. No.</b>	<b>Description of the Figure</b>	<b>Page No.</b>
1.1	PROP-INDEX BENGALURU	
1.2	SALES OF HOUSE (UNIT WISE)	

## NOMENCLATURE USED

Example	
ML	Machine Learning
REGRESSION	Linear Regression
Prediction System	ML model
Accuracy	Metrics
ML Model	Machine learning model

# Chapter 1

## INTRODUCTION

### 1.1. Overview

It describes the overview and the problem definition of the model. Objectives are discussed.

### 1.2. Problem Definition

Machine learning has been used for years to offer image recognition, spam detection, natural speech comprehension, product recommendations, and medical diagnoses. Today, machine learning algorithms can help us enhance cyber security, ensure public safety, and improve medical outcomes. Machine learning systems can also make customer service better and automobiles safer. When we started experimenting with machine learning, we wanted to come up with an application that would solve a real-world problem but would not be too complicated to implement. We also wanted to practice working with regression algorithms. So I started looking for a problem worth solving. Here's what we came up with. If you're going to sell a house, you need to know what price tag to put on it. And a computer algorithm can give you an accurate estimate! With the given features (categorical and continuous) build a model to predict the price of houses in Bengaluru.

### **1.3. Objectives**

As a first project, we intended to make it as instructional as possible by tackling each stage of the machine learning process and attempting to comprehend it well. We have picked Bangalore Real Estate Prediction as a method, which is known as a "toy issue," identifying problems that are not of immediate scientific relevance but are helpful to demonstrate and practice. The objective was to forecast the price of a specific apartment based on market pricing while accounting for various "features" that would be established in the following sections.

### **1.4. Methodology :**

#### **Data Collection**

The statistics were gathered from Bangalore home prices. The information includes many variables such as area type, availability, location, BHK, society, total square feet, bathrooms, and balconies.

Followed by fitting the dataset to the models like:

- 1) Linear Regression
- 2) Lasso regression
- 3) Ridge regressin
- 4) Support Vector Machine
- 5) Random Forest Regressor

## **1.5. Hardware and Software Tools Used**

### **Software:**

- ANACONDA ENVIRONMENT
- JUPYTER NOTEBOOK
- JUPYTER LABS

### **Hardware:**

- ANY LAPTOP/PC WILL BE SUFFICIENT.
- GRAPHIC CARD IS ADVISABLE
- RAM ABOVE 8GB IS ADVISABLE

## **Chapter 2**

### **LITERATURE SURVEY**

#### **2.1. Related Work**

Real Estate Property is not only a person's primary desire, but it also reflects a person's wealth and prestige in today's society. Real estate investment typically appears to be lucrative since property values do not drop in a choppy fashion. Changes in the value of the real estate will have an impact on many home investors, bankers, policymakers, and others. Real estate investing appears to be a tempting option for investors. As a result, anticipating the important estate price is an essential economic indicator. According to the 2011 census, the Asian country ranks second in the world in terms of the number of households, with a total of 24.67 crores. However, previous recessions have demonstrated that real estate costs cannot be seen. The expenses of significant estate property are linked to the state's economic situation. Regardless, we don't have accurate standardized approaches to live the significant estate property values.

First, we looked at different articles and discussions about machine learning for housing price prediction. The title of the article is house price prediction, and it is based on machine learning and neural networks. The publication's description is minimal error and the highest accuracy. The aforementioned title of the paper is Hedonic models based on price data from Belfast infer that submarkets and residential valuation this model is used to identify over a larger spatial scale and implications for the evaluation process related to the selection of comparable evidence and the quality of variables that the values may require. Understanding current developments in house prices and homeownership are the subject of the study. In this article, they utilized a feedback mechanism or social

pandemic that fosters a perception of property as an essential market investment.

## **2.2. Existing System**

**A. User-based approach : N/A**

**B. Item-based approach : N/A**

## **2.3. Limitation of Existing System**

Since we are tabulating the accuracies and displaying the summary of the results. It requires no limitations , as this is just an analysis.

## **2.4. Proposed System**

Nowadays, e-education and e-learning is highly influenced. Everything is shifting from manual to automated systems. The objective of this project is to predict the house prices so as to minimize the problems faced by the customer. The present method is that the customer approaches a real estate agent to manage his/her investments and suggest suitable estates for his investments. But this method is risky as the agent might predict wrong estates and thus leading to loss of the customer's investments. The manual method which is currently used in the

market is out dated and has high risk. So as to overcome this fault, there is a need for an updated and automated system. Data mining algorithms as well as Machine Learning algorithms can be used to help investors to invest in an appropriate estate according to their mentioned requirements. Also the new system will be cost and time efficient .This will have simple operations. In our project, the proposed system works on Linear Regression Algorithm. 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. In this paper, the linear regression algorithm helps to full fill customers by increasing the accuracy of estate choice and reducing the risk of investing in an estate. A lot of features that could be added to make the system more widely acceptable. We have applied very efficient and logical feature extraction techniques so as to increase the accuracy. For example, we have done removal of outliers by using the business logic and bathroom feature. We know that if someone is going to buy a house of say 2000 sqft, then he should have at least 3 to 4 bedrooms. Any other cases in which there are only 2 rooms in 2000 sq ft, has been removed as an outlier. Another feature is the removal of cases where there are absurd numbers of bathrooms. By our logic, the total number of bathrooms should be at most 1 more than total number of bedrooms, i.e.  $\text{total bath} = \text{total BHK} + 1$ . Therefore, we removed any other cases from the data frame where the cases were contradictory. 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.



## Chapter 3

### METHODOLOGY

#### 3.1. Dataset

There's a lot of data involved in fully training the model. The dataset is kept under a same directory. All preprocessing scripts will, by default, output the clean data to a new directory created in the datasets root directory. The following dataset has been used: • Bengaluru\_house\_prices.csv .This dataset was prepared as a record for the house prices of different houses at different locations in the city of Bengaluru by various government authorities. This dataset is a large collection of over 13321 records and 9 columns of house price data collected by various trusted sources. It consists of the following features: area\_type, availability, location, size, society, total\_sqft, bath, balcony, price. In these features, the price column is the labelled attribute.

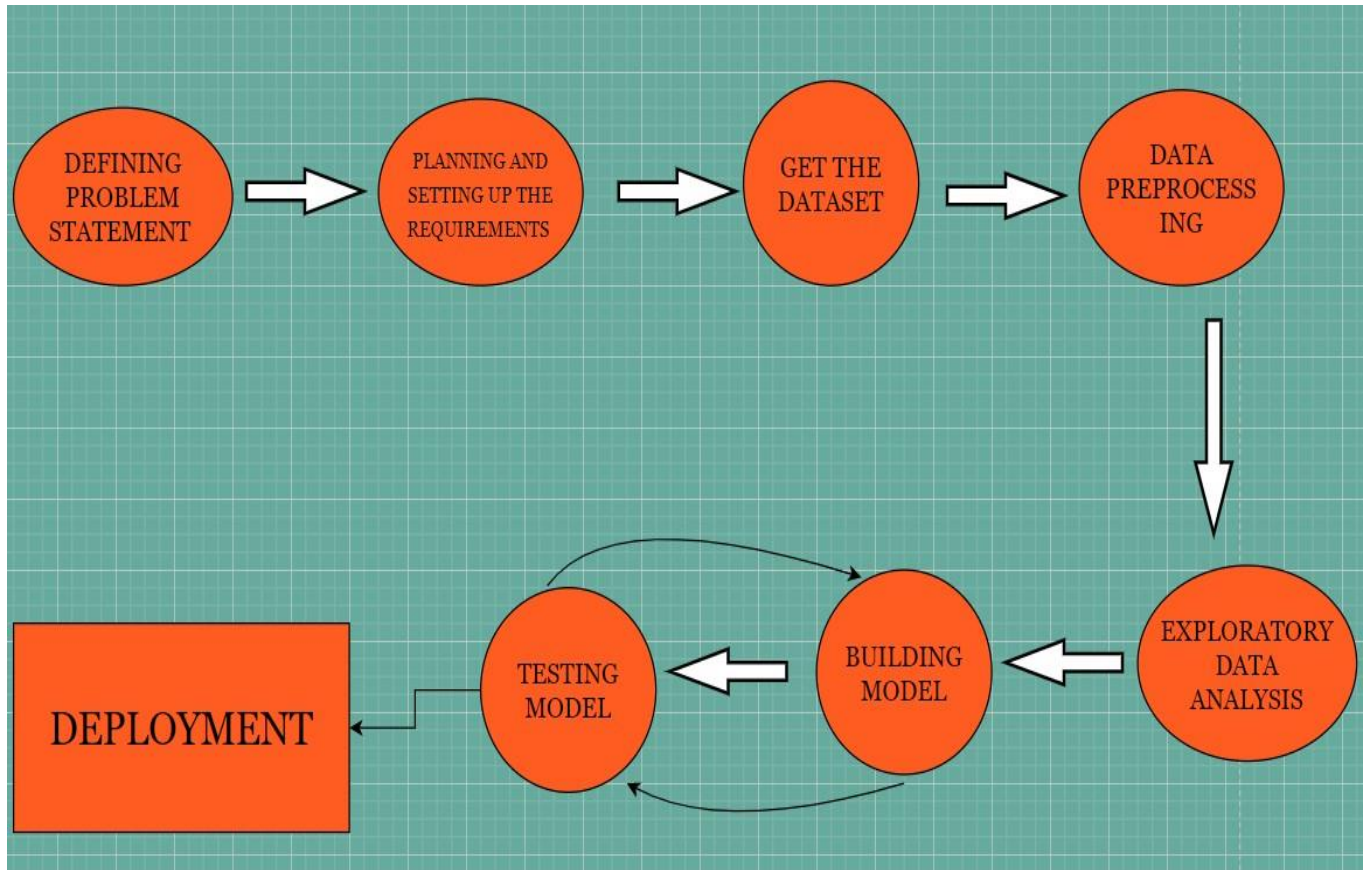
Link:[http://www.kaggle.com/dataset/bengaluru\\_house\\_price/](http://www.kaggle.com/dataset/bengaluru_house_price/)

#### 3.2. Architecture

**Steps involves :**

1. Pre-Processing and Data Cleaning
2. Feature Engineering
3. Dimensionality Reduction and Outlier Removal
4. Model Building and Accuracy
5. Summary

### 3.3. Sequence Diagram



## **Chapter 4**

### **TOOL DESCRIPTION**

This section gives a detailed description about the hardware tools and software tools involved in developing this system and how they are used.

#### **4.1. Hardware Requirements:**

- 1) ANY LAPTOP/PC WILL BE SUFFICIENT.
- 2) GRAPHIC CARD IS ADVISABLE
- 3) RAM ABOVE 8GB IS ADVISABLE

#### **Software Requirements :**

- ANACONDA ENVIRONMENT
- JUPYTER NOTEBOOK
- JUPYTER LABS

## **Chapter 5**

### **IMPLEMENTATION**

#### **STEPS TO RUN THE PROGRAM:**

1. Import Libraries
2. Load Dataset
3. Exploratory Data Analysis
4. Data Cleaning
5. Feature Engineering
6. Dimensionality Reductions
7. Outlier Removal using Business Logic
8. Outlier Removal using Standard Deviation & Mean
9. Data Visualization
10. Building a Model
11. Test the Model for few properties
12. Export the tested model to a pickle file



## Chapter 6

### RESULTS AND ANALYSIS

This section is containing a description about the main findings of this proposed system with the figures, detailed explanation by comparing with previous studies and analysis.

#### 6.1. Result Discussion

Model	Score
LinearRegression	0.790384
Lasso	0.790384
Ridge	0.803637
Support Vector Machine	0.20638
Random forest Regressor	0.884665

#### 6.2. Analysis

Use of ridge and random forest regressor is ideal according to results. Use of support vector machine is ruled out .

## **Chapter 7**

### **CONCLUSIONS AND FUTURE SCOPE**

The framework makes ideal utilization of the Linear Regression Algorithm. It makes use of such information in the most effective way. The direct relapse calculation satisfies. customer by expanding the exactness of their decision and diminishing the danger of putting resources into a home. One of the real future extensions is including home database of more urban areas which will give the client to investigate more domains and achieve an exact choice. More factors like subsidence that influence the house costs should be included. Top to bottom subtle elements of each property will be added to give plentiful points of interest of a coveted domain. The authors were able to create a system with more than 85% accuracy and the utilization of dataset was done with great efficiency which ultimately gave quite impressive results.

## REFERENCES

- [1] R Manjula, Shubham Jain, Sharad Srivastava and Pranav Rajiv Kher, —Real estate value prediction using multivariate regression models,|| IOP Conference Series: Materials Science and Engineering, 2017.
- [2] Eduard Hromada, —Mapping of real estate prices using data mining techniques,|| Czech Technical University, Czech Republic, 2015
- [3] Adyan Nur Alfiyatin and Ruth Ema Febrita, —Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization,|| International Journal of Advanced Computer Science and Applications, 2017
- [4] Li Li and Kai-Hsuan Chu, —Prediction of Real Estate Price Variation Based on Economic Parameters,|| Department of Financial Management, Business School, Nankai University, 2017.
- [5] Nissan Pow, Emil Janulewicz and Liu Dave, —Applied Machine Learning Project 4 Prediction of real estate property prices in Montreal,|| 2016.
- [6] Aminah Md Yusof and Syuhaida Ismail, Multiple Regressions in Analysing House Price Variations. IBIMA Publishing Communications of the IBIMA Vol. 2012 (2012), Article ID 383101, 9 pages DOI: 10.5171/2012.383101.
- [7] Babyak, M. A. What you see may not be what you get: A brief, nontechnical introduction to over fitting regressiontype models. Psychosomatic Medicine, 66(3), 411-421.
- [8] Vasilios Plakandaras and Theophilos, Rangan Gupta\*, Periklis Gogas —Forecasting the U.S. Real House Price Index





## APPENDIX - I

### **Example**

#### **SOURCE CODE**

**GitHub:** (link will be uploaded shortly, after implementing all the censorship and privacy measures.)

**Note:** This model is still in development phase and shall be proprietary and may be

commercialized if the creators wish to do it.