

Real Estate Price Prediction

Govind Kumar, Priyanka Makkar, Dr Yojna Arora

ABSTRACT- Analysing various fields, associate numbers, events became the need of time and most important step to do anything and hence data science became an important part in every field. Using the concept of data science, the project of Real Estate Price Prediction is built. The motive of creating a project on Real Estate Price Prediction was just to implement the concepts of data science and python language that is used in analysing for designing an application. This was done to get better understanding of the skills that are needed in python language, analysis using data science. The project focuses on the different features and algorithm available in python and data science. In this project various library of python is used to design an attractive, effective, and beautiful project. The project will introduce a Real Estate price estimation system that done estimate based on various mathematical algorithms and tricks and then gives best possible result. So basically, what this application does is identify the need of user in any specific area in Bangalore.

KEYWORDS- Data science, Analysis, Price, Estimation, Bangalore

I. INTRODUCTION

In the business of Real Estate there are a lot of sellers who is selling their property in that area and if the buyer has a limited budget for spending, so one must have to do a lot of research to meet his requirements. So here the need of data science occurs as with the help of analysis one can get an approximate cost and general idea of price and availability of the property.

There is a lot of difference in doing analysis and deep & good analysis. Analysis is just an overview of the records

whereas deep analysis gives us much better result by analysing peak, average factors alone with outlier values etc. As it is supposed that outliers affect the mean result, so all algorithmic calculation are done to overcome that inconsistency in result. This gives user a best result to do things and hence this saves the time of the user and benefits him a lot in searching desired result only. Also, in some cases sometimes seller put cost of his property so high. So, after analysis of data an average estimation is provided that might help the user from fraud type activities and false result. In this project database is taken is of Bangalore city. Algorithms made calculation on data present in database only [1] [2].

II. RELATED WORK

Data science helps to evaluate data more deeply, logically and in scientific manner which helps the user in various fields to study weak and strong points if considered in field of business or research or invention or discovery or anything done to make the world a better place or any activities done by one for his own benefits and need. When considered in field of decision-making analysis plays a vital role here also as it analyses all the implementation and after proper logical and scientific evaluation it provides best results. Analysing must be done before doing any task because it decreases the risk of failure and provides possible fact and figures where work should be done because while doing any task or job there are various sections where individual or group or organisation have to focus but its became hard for them to identify the particular part. As if this project is considered there are a lot of sellers of various kind of properties in a particular area and it is very hard for anyone to go through all the records and then select because it requires a lot of efforts and power and overall it is actually the wastage of money and considered as bad or worst approach of doing a task [3] [4].

Manuscript received October 10, 2020

Govind Kumar, Department of Computer Science & Engineering, Amity University, Haryana, India (govinddgr@gmail.com)

Ms Priyanka Makkar, Assistant Professor, Department of Computer Science & Engineering, Amity University, Haryana, India

Dr Yojna Arora, Assistant Professor, Department of Computer Science & Engineering, Amity University, Haryana, India

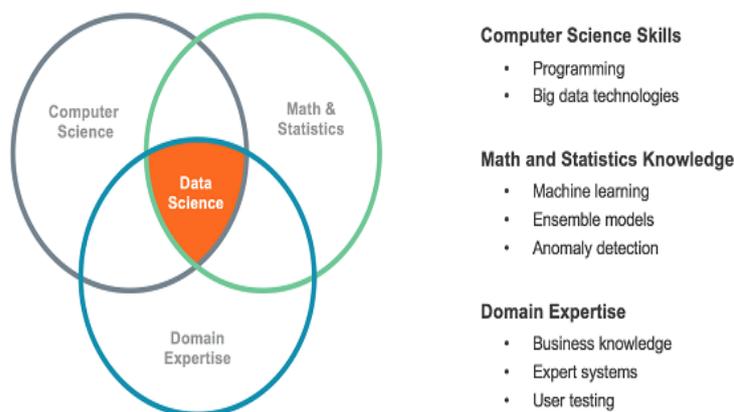


Fig 1: Data Science → A powerful Combination of various disciplines

Figure 1 above is showing that Data Science is the combination of various disciplines that are very important key factors in any field. Each field is very important from various perspectives while analysing best possible result.

III. METHODOLOGY

While doing python code for this project various steps were followed to analyse the data in best possible way so that there will be no contractionary outputs occurs while retrieving the data.

Some steps followed while making the project are listed below:

- 1) Read csv file.
- 2) Data Load: Load Bangalore home prices into a data frame.
- 3) Data Cleaning: Handle NA values.
- 4) Feature Engineering.
- 5) Add new feature(integer) for BHK (Bedrooms Hall Kitchen).
- 6) Explore total_sqft feature.
- 7) Add new feature called price per square feet.
- 8) Examine locations which is a categorical variable.
- 9) Dimensionality Reduction: Any location having less than 10 data points should be tagged as "other" location.
- 10) Outlier Removal Using Business Logic.
- 11) Outlier Removal Using Standard Deviation and Mean.
- 12) Plot same scatter chart to visualize data.
- 13) Use One Hot Encoding for Location.
- 14) Use K Fold cross validation to measure accuracy of our Linear Regression model.
- 15) Find best model using GridSearchCV.
- 16) Test the model for few properties.
- 17) Export the tested model to a pickle file.

- 18) Export location and column information to a file that will be useful later on in our prediction application.



Fig 2: Data Science Workflow

Shortlisting Properties and Property Valuation is done in real estate program which can help user to shortlist the most suitable properties that meet his requirement. Suppose user have an elderly person in the family who requires frequent medical attention, and user have no time to visit each property then this program will help him. Property Valuation is one of the important things that user needs when buying or selling real estate is a price estimation. The project uses best possible algorithms to analyse the cost of the property and helps the user a lot[5].

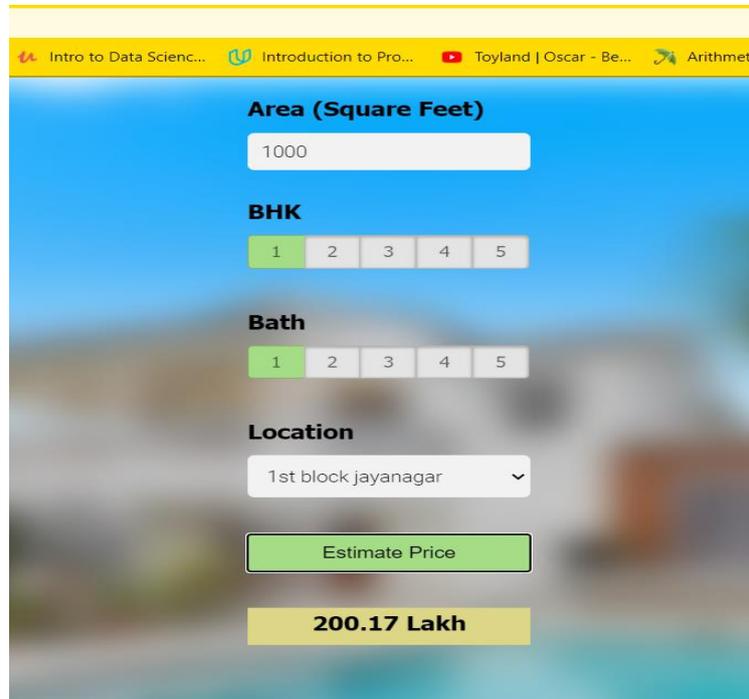
IV. RESULT

The result here is that the project developed by the author gives the best price estimation. The estimation of price is done in steps i.e. at first shortlisting of property is done as per user requirements and then cost of shortlisted properties is evaluated. Various algorithms like outliers, K Fold cross validation etc is done to get best analysis even though there is uncertainty in few data values. This will let user in better position to evaluate the actual

market value of the apartment, house, land, or whatever user wish to buy. The project gives the brief introduction that how analysis is done and how algorithms can be implemented in better and easy way[6].

As it is clearly visible in the snapshot Figure 3 that user is asked to input various fields like Area, BHK, Bath and

must select the location from given options. Then after getting all user requirements the program will shortlist that property from database then, evaluate the cost of those properties and then let the user know about the price estimation.[8][9]



The screenshot shows a web application interface for price estimation. It features a blue header with a yellow bar at the top. Below the header, there are several input fields: 'Area (Square Feet)' with a text input containing '1000', 'BHK' with a row of buttons (1, 2, 3, 4, 5) where '1' is selected, 'Bath' with a row of buttons (1, 2, 3, 4, 5) where '1' is selected, and 'Location' with a dropdown menu showing '1st block jayanagar'. A green button labeled 'Estimate Price' is positioned below the location field. At the bottom, a yellow box displays the estimated price: '200.17 Lakh'.

Fig 3: Price estimation project snapshot.

V. FUTURE SCOPE AND CONCLUSION

Through this project the author wants to state that 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. One of the major future scopes is adding database of more cities which will provide the user to explore more estates and reach an accurate decision. 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. While doing this project author used python coding because coding in Python is comparatively very easy to do as Python have some predefined functions and it that makes the task for users quite easy and also reduces the length of the code. In the Real Estate project analytical algorithms and methods are used because that is what data science really is. Data science is the most important concept to learn from various perspective and designing a new model.

REFERENCES

- [1] Aaron Ng, 2015, Machine Learning for a London Housing Price Prediction Mobile Application
- [2] <https://www.dataquest.io/blog/what-is-data-science/>
- [3] Manny Bernabe, expertise in research and deployment of quantitative strategies in exchange traded funds (ETFs).
- [4] <https://insidebigdata.com/2017/07/27/defining-data-science-landscape/>
- [5] Vishal Raman, May 2014. Identifying Customer Interest in Real Estate Using Data Mining.
- [6] <http://www.99acres.com/property-rates-and-pricetrends-in-bangalore>
- [7] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, 2015. Introduction to Linear Regression Analysis
- [8] Gongzhu Hu, Jinping Wang, and WenyingFengMultivariate Regression Modelling for Home Value Estimates with Evaluation using Maximum Information Coefficient
- [9] Iain Pardoe, 2008, Modelling Home Prices Using Realtor Data