House Price Prediction Using Machine Learning

Harshali S. Zalte[1], Prachi S. Patil[2], Prajakta P. Chaudhari[3], Kaustubh R. Kolhe[4],

Jatin S. Bornare[5]

[1-5]UG Student, Computer Engineering, Shram Sadhana Trust's, College Of Engineering and Technology, Jalgaon, India

Abstract: House price forecasting is an important topic of real estate. Machine learning techniques are applied to analyze historical property transactions to discover useful models for house buyers and sellers. Revealed is the high discrepancy between house prices in the most expensive and most affordable suburbs in the area. Significant time and expertise are needed to customize the model for a specific problem. A significant way to reduce the complicated design is by using Automated Machine Learning that can intelligently optimize the best pipeline suitable for a problem or dataset.

Keywords: House Price Prediction, Machine Learning, Google Oauth, Python.

I. INTRODUCTION

Proposed system aims to make a machine learning model which could predict the real estate house prices. Respective chapter has in view of the content focusing on scope, motivation, objectives and selection of life cycle model of the projected system. It also states the actual problem definition which is to be solved using machine learning models.

II. PROJECT PLANNING AND MANAGEMENT.

Requirements analysis is a software engineering task that has bridged the gap between system level requirements engineering and software design requirement engineering activities resulting in the specification of software's operational characteristics (functional, data, and behavioral), that indicates software's interface with other system elements and establishes constraints that software must meet. Gathering of requirements related to the proposed system is done. Analysis is done by comparing these requirements with working and function provided by existing forum system.

A]   Feasibility Study: While considering the feasibility of any system, the system must meet its best performance by three sides i.e., technical, operational and economical feasibility. Feasibility should be considered in any organization since it helps in selection of best system for the job. It is necessary to evaluate the feasibility of the project at the early stages of software development. If project risk is great, the feasibility of product is reduced. A well-designed feasibility study should provide a description of the product or service, details of operation and working of system.

B]   Risk Analysis: Risk analysis and management are actions that help a software team to understand and manage uncertainty. A risk is a potential problem it might happen, or it may not. Risk is a uncertain event or condition that, if it occurs, has a positive or negative effect on a project's objectives. Risk always involves two characteristics: uncertainty- the risk may or may not happen; that is, there are no 100 percent probable risk and loss- if the risk becomes a reality, unwanted consequences or losses will occur. When risks are analyzed, it is important to quantify the level of uncertainty and the degree of loss associated with each risk.

III. SOFTWARE REQUIREMENTS SPECIFICATION.

A]   Product Features:
•   The interface of the application should be user friendly.
•   Accurately predict the house prices, which failing to predict can lead to the project failure.
•   The project should be easy-to-use for the end user. [1]
B]   Operating Environment:
•   Operating System: Windows/Linux/MacOS.
•   Software: Internet Browser.
•   Hardware: PC or Laptop, Server.
C]   Assumption:
•   Dataset Opted to work on should be outlier free with no blank values.
•   Machine learning model should not be over-fitted on the training data, which can reduce the precision of

machine learning model.

IV.   SYSTEM DESIGN

Design is a meaningful engineering representation of something that is to be built. It can be traced to system requirements and at the same time assessed for quality against a set of predefined criteria for "good" design. Software design sits at the technical kernel of software engineering and is applied regardless of software process model that is used. Software design is the first of three technical activities-design, code generation and test that are required to build and verify the software.

A] System Architecture: Figure below is Architecture of Proposed System shows actual working flow of proposed system. The architecture is quite easy to understand for a new user. Authentication and house price prediction are the two main modules of the architecture. Authentication uses API's to authenticate the user and house price prediction uses the machine learning algorithm to predict prices.
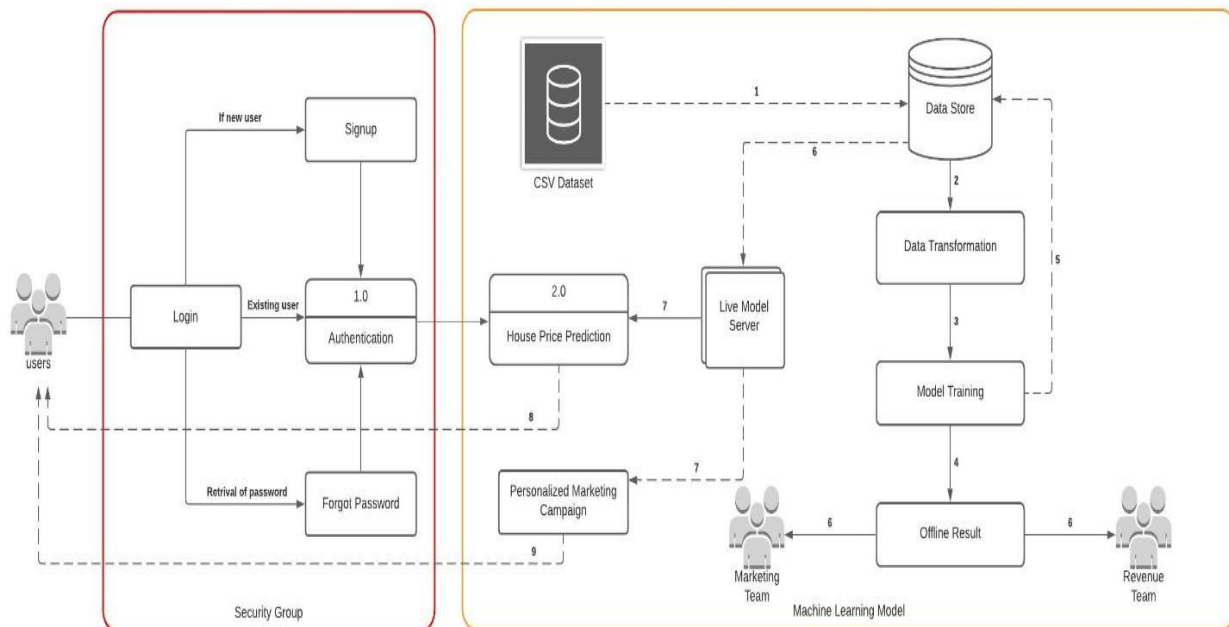
V.   IMPLEMENTATION

A]   Algorithm:
•   Step 1: START
•   Step 2: User Login through Google.
•   Step 3: Authenticating and conforming credentials through API.
•   Step 4: User inputs required Information.
•   Step 5: The Information is then sent to the Backend.
•   Step 6: The Information obtained is processed by trained Machine Learning Model.
•   Step 7: The Model prepares an answer as output.
•   Step 8: Output is displayed to the User.
•   Step 9: STOP

VI.   RESULT

After training the dataset on three different machine learning model the outcome that has been extracted is as follows, the linear regression model performed the best with the score of approximately 91%, followed by the lasso regression with an approximate score of 91%, lastly the decision tree score is almost 90% score when trained on a dataset.

As the final decision to choose over these machine learning models the optimal choice had to be the linear regression model and hence, that is the reason of using the same in the proposed solution. Following figure shows Final Results of Machine Learning model algorithms.

Model	Scores in %	Best Parameters
Linear Regression	90.686856	{'normalize': True}
Lasso Regression	90.684626	{'alpha': 2, 'selection': 'cyclic'}
Decision Tree	89.400316	{'criterion': 'mse', 'splitter': 'random'}

VII. 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 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.

VIII. REFERENCES

[1]   Edna Dias Canedo and Bruno Cordeiro Mendes. Regression Machine Learning. Department of Computer Science, University of Brasília (UnB), P.O. Box 4466, Brasília 70910-900, Brazil.

[2]   Atharva Chogle, Priyanka Khaire, Akshata Gaud and Jinal Jain4 House Price Forecasting. [House price forecasting using machine learning]. Bachelor of Engineering, Dept. of Computer Engineering, RGIT, Mumbai, Maharashtra, India.

[3]   Suraya Masrom, Thuraiya Mohd, Nur Syafiqah Jamil and Norhayati Baharun Automated Machine Learning. IEEE 2020.

IX.   BIOGRAPHIES

Ms. Harshali S. Zalte:

•   **Qualification:** UG candidate of Computer Engineering, Shram Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon.

•   **Area of Interests:** HTML, CSS, JavaScript, MySQL, Algorithms.

Ms. Prachi S. Patil:

•   **Qualification:** UG candidate of Computer Engineering, Shram Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon.

•   **Area of Interests:** C, Python, Java-1, Project Management, Ethical Hacking.

Ms. Prajakta P. Chaudhari:

•   **Qualification:** UG candidate of Computer Engineering, Shram Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon.

•   **Area of Interests:** C, C++, Java, PHP, Software testing.

Mr. Kaustubh R. Kolhe:

• **Qualification:** UG candidate of Computer Engineering, Shram Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon.

• **Area of Interests:** Python, AI, Data Handling, Database Management.

Mr. Jatin S. Bornare:

• **Qualification:** UG candidate of Computer Engineering, Shram Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon.

• **Area of Interests:** Python, AI, Machine Learning, Flask, MySQL, GCP, Data analysis.