ABSTRACT

The real estate market is one of the most competitive in terms of prices and it tends to vary significantly based on a lot of factors, hence it becomes one of the prime fields to apply the concepts of machine learning. House prices are a significant impression of the economy, and its value ranges are of great concerns for the clients and property dealers. Property prices escalate every year that eventually reinforced the need for strategy or technique that could predict house prices in future. There are certain factors that influence house prices including total size, locations, number of bedrooms and others. Traditionally predictions are made on the basis of these factors. However such prediction methods require appropriate knowledge and experience regarding this domain. Machine Learning techniques have been a significant source of advanced opportunities to analyse, predict and visualise housing prices. Here Regression models such as Linear Regression, Lasso Regression and Decision Tree are utilised to predict property prices.

I. INTRODUCTION

The real estate market is a complex and dynamic industry, with house prices varying widely depending on location, size, and other relevant features. Predicting house prices accurately can be a challenging task, but with the increasing availability of data and the advancements in machine learning techniques, it is now possible to develop models that can predict house prices with high levels of accuracy. In this project, we aim to develop a web-based tool that uses machine learning techniques to predict house prices.

The tool will take input from the user, such as the location, size, number of bedrooms, and other relevant features of the house, and use a trained machine learning model to predict its price. The project will involve collecting and preprocessing data, selecting the appropriate machine learning algorithm, and training the model. We will then integrate the model into a web-based user interface, allowing users to easily input data and obtain house price predictions. The target audience for this tool includes real estate agents, homebuyers, and homeowners looking to sell their property. The tool will be designed to be user-friendly and accessible, even for those without a background in data science or machine learning. Overall, this project aims to provide a useful and accessible tool for predicting house prices, allowing users to make more informed decisions in the real estate market.

II. SCOPE

The main scope of the project is to predict accurate prices of houses and develop a web-based tool that allows users to predict house prices using machine learning techniques.

The tool will take input from the user, such as the location, size, number of bedrooms, and other relevant features of the house, and use a trained machine learning model to predict its price. The project will involve collecting and preprocessing data, selecting the appropriate machine learning algorithm, and training the model. We will then integrate the model into a web-based user interface, allowing users to easily input data and obtain house price predictions. Overall, the scope of this project is to provide a useful and accessible tool for predicting house prices, allowing users to make more informed decisions in the real estate market.

III. PROBLEM DEFINITION

Prices of real estate properties are sophisticatedly linked with our economy. Let’s say we are a real estate agent, and we are in charge of selling a new house.

We don’t know the price, and we want to infer it by comparing it with other houses. We look at features of the house which could influence the house, such as size, number of rooms, availability, bathrooms, area type, etc. At the end of the day, what we want is a formula on all these features which gives us the price of the house, or at least an estimate for it. The goal of our project is to use machine learning algorithms such as Linear Regression, Lasso Regression and Decision Tree and predict the selling prices of property such as houses.
IV. LITERATURE SURVEY

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 standardised 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 utilised a feedback mechanism or social pandemic that fosters a perception of property as an essential market investment.

V. OBJECTIVE

The objective of this project is to develop a web-based tool that uses machine learning techniques to accurately predict house prices. The tool will be designed to be user-friendly and accessible, allowing users to obtain reliable house price predictions quickly and easily. The specific objectives of the study are:

❖ To apply data preprocessing and preparation techniques in order to obtain clean data
❖ To build machine learning models able to predict house price based on house features
❖ To analyse and compare models performance in order to choose the best model
❖ To develop a web-based tool that is user-friendly and accessible, allowing users to obtain reliable house price predictions quickly and easily.

VI. METHODOLOGY

1. Pre-Processing and Data Cleaning

Data preprocessing is an integral step in Machine Learning as the quality of data and the useful information that can be derived from it directly affects the ability of our model to learn; therefore, it is extremely important that we preprocess our data before feeding it into our model. Feature Engineering

2. Feature Engineering Feature engineering

It is the process of using domain knowledge of the data to create features that make machine learning algorithms work. If feature engineering is done correctly, it increases the predictive power of machine learning algorithms by creating features from raw data that help facilitate the machine learning process. Feature Engineering is an art. In our project, it includes exploring the total_sqft feature and also adds new feature price per square feet.

3. Dimensionality Reduction and Outlier Removal

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. In our project, any location which had a number of houses less than 10 has been marked as “others” so as to reduce the dimensions of the dataset. Outliers badly affect mean and standard deviation of the dataset. These may statistically give erroneous results. It increases the error variance and reduces the power of statistical tests. If the outliers are non-randomly distributed, they can decrease normality. So we applied various logics such as business logic, bathroom features to remove the outliers.
4. Model Building and Accuracy –
In our project, the model was implemented using the Linear Regression Algorithm. All the necessary libraries were imported and training of the model was done. We saw that in 5 iterations we get a score above 85% all the time. This was a very good accuracy score and we continued to use the algorithm. Also we compared different algorithms, such as Lasso Regression, Decision Tree and Linear Regression using the GridSearchCV technique to find the model with best accuracy, which we found to be Linear Regression.

VII. PROJECT ARCHITECTURE

Architecture of the Application

VIII. EXPERIMENTAL SETUP

Steps to Create Model:
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
13. Deploy the model as a web page based tool

Technologies used:
1. Python
2. HTML
3. CSS
4. Bootstrap
IX. RESULTS

In conclusion, we used machine learning techniques to predict house prices. We collected a dataset and preprocessed it to ensure it was ready for analysis. We then used various machine learning algorithms, such as linear regression, decision trees and lasso regression, to train models to predict house prices. After evaluating the performance of these models, we selected the most accurate model and optimised its hyperparameters. We then deployed the model to make predictions on new data and showed how it can be used in practice. Our results demonstrate that machine learning techniques can be powerful tools for predicting house prices, with high levels of accuracy and precision. However, we also acknowledge that there are limitations to this approach, such as the need for high-quality data, potential for overfitting, and the need to incorporate domain knowledge. Future research could explore the use of more advanced machine learning techniques, such as deep learning, or the integration of additional data sources to improve the accuracy and usefulness of these models. Overall, our work highlights the potential of machine learning in real estate and housing markets, and underscores the need for continued innovation in this field.

X. CONCLUSION

XI. REFERENCES