International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 536-541 www.ijaem.net ISSN: 2395-5252

Web Application on House Sale Prediction

Mrs. Shalani Kushwaha, Aanchal Kumari, Arpit Kushwaha

Date of Submission: 03-02-2023 Date of Acceptance: 17-02-2023

ABSTRACT
The purpose of this paper is to predict the market value of properties for sale. This program helps find the starting price of a place based on the variables of the place. Similarly, consider a situation where a person needs to sell a house. So, in both cases, you canbe confident that the price of the home is good for both the buyer and the seller. Realestate prices are rising year by year, and there is a need for a real estate forecasting system. Estimating the value of a homehelps developersdetermine how much a home will sell forand helps customers determine whenitis reasonable to purchase a home.
Research shows that home price volatility is often a problem for homeowners and estate market. A literature review is conducted to analyse relevant properties and the most efficient model compared to other models. Furthermore, our results also suggest that location and structural attributes are important factors in predicting house prices. The study is particularly useful for property developers and researchers to identify the most important attributes used to determine property prices.
Keywords: Linear Regression Models,Python, Machine Learning.

I. INTRODUCTION

Machine learning has played a big role in choosing images, spamming, and sharing voice commands for years. Machine learning also provides better customer service and safer vehicle systems. The housing market stands out among the most price-sensitive markets and is constantly evolving. This is one of the major areas where machine learning ideas are applied to improve and predict costs with high accuracy.

The housing market has positive impact on the national currency, which is the size of an important national economy. Homeowners purchase goods such as furniture and appliances for their homes, and 4,444 homebuilders or contractors purchase raw materials to build their homes to meet housing demand. In a metropolitan area like Bangalore, potential buyers consider several factors such as location, size, proximity to parks, schools, hospitals, power stations and most importantly, home value increases. Machine learning is database system learning process. Machine learning is the part of data science that uses machine learning algorithms to process data. Supervised strategy is a hypothetical model used for activities involving predicting values form other values in a dataset. Supervised reading has predefined labels. Separate objects based on the parameters of one of the predefined labels. Supervised learning methods descriptive models are used in the activities that benefits from the understanding gained by summarizing data in new and exciting ways. The purpose of this statistical analysis is to help understand the relationship between home characteristics and how these terms are used to predict home prices.

1.1 MACHINE LEARNING

Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data. Classical machine learning is categorical and is often categorized by how an algorithm learns to become more accurate in its predictions through analysis of previous and current data information. There are four basic approaches of machine learning and it is the supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning.Machine learning is important because it gives enterprises a view of trends in customer behaviours and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.[2]

1.2 WEB APPLICATION

A typical flow for a web application looks like this:

- A user initiates a request to her web server over the internet, either through a web browser or through her user interface of an application. Web server forwards this request to the appropriate web application server.
- Web application servers perform request tasks such as querying databases, processing data,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 536-541 www.ijaem.net ISSN: 2395-5252

and producing results for the requested data.[2]

- Web application servers generate results contains requested information or processed data to the web server.
- The web server responds the requested information is returned to the client and displayed on the user's display.

II. METHODOLOGY

The methodology for a house sale prediction using linear regression could include the following steps:

1. Collection of Data: gain a dataset of housing income, which includes applicable variables consisting of the sale price, size, place, age, variety of bedrooms, and number of bathrooms.
2. Information Pre-processing: smooth and pre-process the facts to make certain that it's far suitable for evaluation. this can contain filling in lacking values, remodelling variables to fulfil assumptions of linear regression, and getting rid of outliers.
3. Variable choice: choose the independent variables to be used within the linear regression model primarily based at the statistics available and any prior knowledge of the housing marketplace.
4. Model building: build a linear regression version using the chosen independent variables and the sale fee as the based variable. check the assumptions of linear regression, together with linearity, independence, homoscedasticity, and normality.
5. Version evaluation: compare the performance of the model using statistical measures together with R-squared, suggest absolute error, and mean squared mistakes. examine the performance of the version with other fashions or benchmarks.
6. Version Refinement: Refine the model if necessary, by way of including or getting rid of variables, transforming variables, or the usage of a distinctive regression method.
7. Predictions: Use the final version to make predictions on new facts, which includes the sale fee of a specific assets.
8. Conclusion: Draw conclusions approximately the performance of the version, the significance of the independent variables, and the capacity of the version to accurately predict housing sale costs.

III. EXPERIMENTAL SETUP

The solution to the problem is to reset the line. This process models the relationship between the target variable and the independent variables (predictors). Identifies the line model with the coefficients in the data to reduce the remaining squared amount between the target variation in the data and the predicted value from the line measurement. This record has many features such as:

- Imported libraries like pandas, NumPy, Matplotlib, Sklearn, Seaborn, Warnings. Then treated outliers, and then treated missing values.
- Afterwards we use the type conversations if needed for any features. The next process is used is to label encoding if needed like we used in 'no of times visited' feature of our dataset.
- Feature generation of the feature 'ever_renovated' is done. Binning for the feature 'Zipcode' is done as it is of integer type but less likable to be integer type. So, we convert feature 'Zipcode' to string type first and then we binned it using bins of 10 as according to feature 'Sale_Price' and made a new feature 'Zipcode_sale'.
- After that we deleted useless variables like 'Zipcode'. Then we Classify target and independent variable.
- Use the techniques to Scale out the data using scaler function of sklearn. Next is to remove multicollinearity using vif (Variance Inflation Factor).
- After we use machine learning model Linear regression model from 'Sklearn.linearmodels' to predict the output and test our model. Then the score of the model is calculated. Later on, we check residuals distribution by plotting. (Our model produces the bell-shaped distribution curve which shows the data is well modeled).
- After that we use flask for implementing ML algorithm in Web app. Flask is a web framework that provides libraries to build lightweight web applications in python.

IV. LINEAR REGRESSION

Linear Regression is a statistical method for modelling the connection between a dependent variable and one or greater unbiased variables. The purpose of linear regression is to find the exceptional-fitting line that describes the connection between the independent variables and the dependent variable. This line is used to make predictions about the dependent variable given new values of the independent variables.[3]

Mathematically, the linear regression can be represented as:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 536-541 www.ijaem.net ISSN: 2395-5252

y= a0+a1x+e
where,
Y=dependentVariable
X=independent Variable

a0=intercept of the line (or it is the additional degree of freedom)
a1=Linear regression coefficient (scale factor to each input value)
e= random error

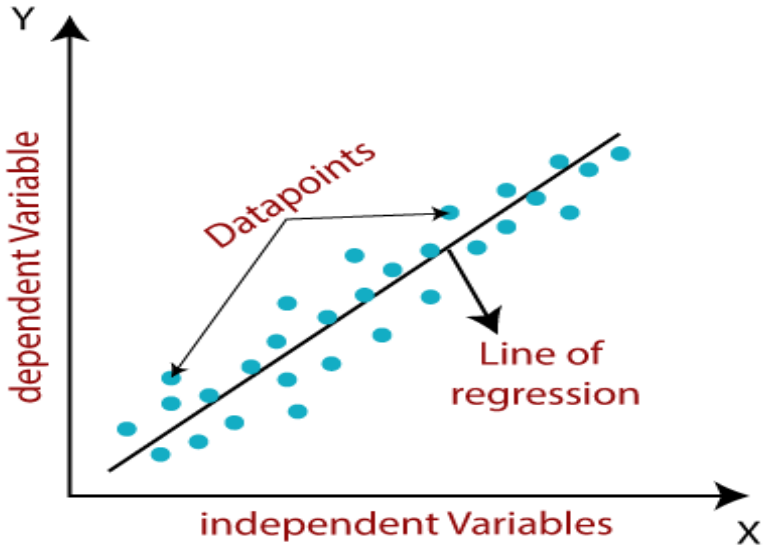

Fig1: Linear Regression model[9]

Linear regression is primarily based on the idea that the relationship among the established and unbiased variables is linear, which means that the alternate within the structured variable is proportional to the exchange in the impartial variables. The fine-becoming line is determined through minimizing the difference among the found values and the predicted values of the structured variable.

Linear regression is broadly used in many fields, consisting of economics, finance, and engineering, for predictive modelling, hypothesis testing, and feature choice. however, it's miles important to cautiously evaluate the assumptions of linear regression, inclusive of linearity, independence, homoscedasticity, and normality, before applying it to a particular dataset.[4]

A linear line showing the relationship between dependent and independent variables is called a regression line. A regression line can show two types of relationships:

- Positive linear relationship:

If the dependent variable increases on the Y-axis and the independent variable increases on the X-axis, then such a relationship is called a positive linear relationship.
Equation for the slope line: Y=A0+A1+X

- Negative linear relationship:

If the dependent variable on the Y-axis decreases and the independent variable on the X-axis increases, then such a relationship is called a negative linear relationship.
Equation for the slope line: Y=A0+A1+X

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 536-541 www.ijaem.net ISSN: 2395-5252

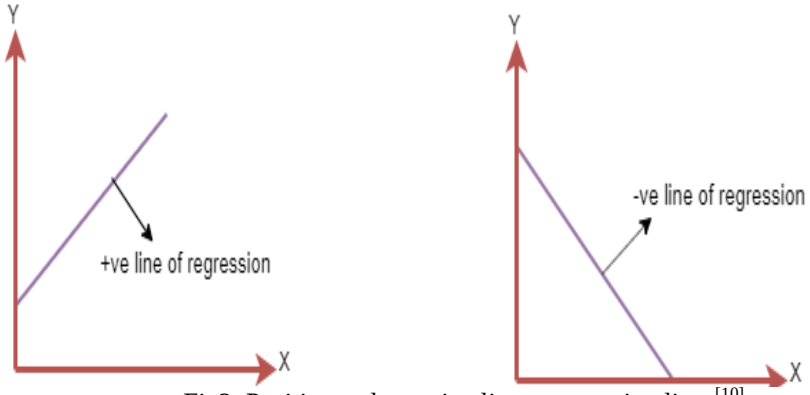

Fig2: Positive and negative linear regression lines[10]

4.1 The Prequisites Used For Linear Regression

Below are some important assumptions of linear regression. These are some formal checks when building a linear regression model that ensures the best possible result is obtained from a given set of data.

Linear regression assumes a linear relationship between the dependent and independent variables. Multicollinearity means high correlation between independent variables. Because of multicollinearity, it may be difficult to find a true relationship between predictors and target variables. Or we can say that it is difficult to determine which predictor variable affects the target variable and which does not. Thus, the model assumes either little or no multicollinearity between traits or independent variables.

Homoscedasticity is when the error term is the same for all values of the independent variables. In the case of homoscedasticity, there should not be a clear distribution of data in the scatterplot.

Linear regression assumes that the error term should follow a normal distribution pattern. If the error terms are not normally distributed, then the confidence intervals become either too wide or too narrow, which can cause difficulty in finding the coefficients.

This can be checked using a q-q plot. If the graph shows a straight line without any deviation, it means that the error is normally distributed.The linear regression model assumes no autocorrelation in the error terms. If there will be any correlation in the error term, then it will drastically reduce the accuracy of the model. Autocorrelation usually occurs when there is dependence between the residual errors.

V. PROJECT IMPACTS ON THE HOUSE SALE PREDICTION SYSTEM

The Home Sales Forecast Linear Regression Model can be used to estimate the expected sales price of a property based on certain factors such as location, size, and number of bedrooms. When you build such a model, you get a formula that describes the relationship between the input variables (predictors) and the output variables (sales price). This model can then be used to make predictions by inputting new data into the equation and obtaining an estimated selling price. The accuracy of a model's predictions depends on several factors, including the quality and quantity of data used to train the model, the strength of the relationship between the predictor and output variables, and the validity of the assumptions made. increase. data.
Contributions are:

- Increase Efficiency: The system automates and streamlines the sales priceforecasting process, reducing the time andeffort required for manual forecasting.
- Better PricingStrategies: The system provides insight into the factors thatinfluence sales prices, enabling realtors and owners to developeffective pricing strategies.
- Market analysis: This system analyses historical sales data and can be used to identify trends and patterns in the housing market, providing valuable insight to real estate industry stakeholders.
- Risk Management: This system can be used to assess the potential risks and uncertainties associated with buying or selling real estate, helping investors and property owners make informed investment decisions. Helpful. Overall, the Home Sales Forecasting System helps increase the efficiency and effectiveness

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 536-541 www.ijaem.net ISSN: 2395-5252

of the real estate market, benefiting all involved.

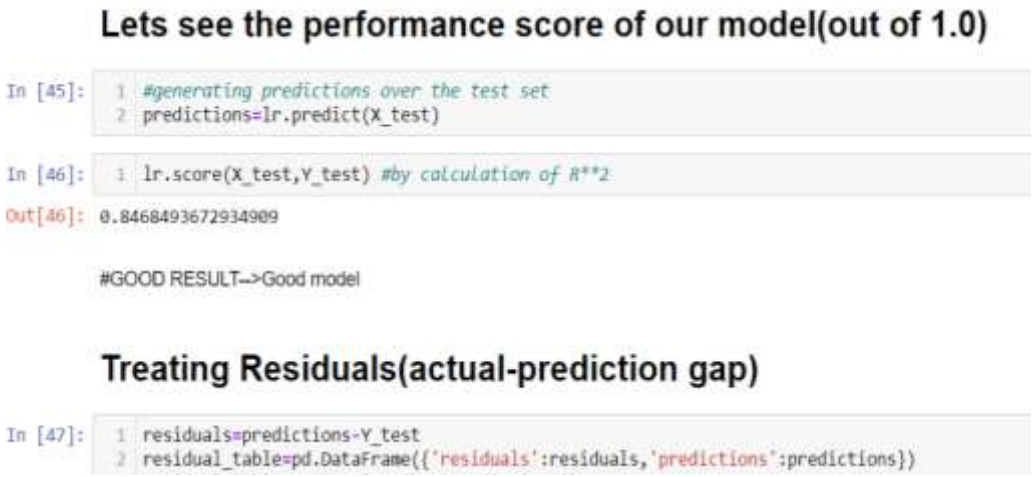

Fig3: Performance of linear model of the project

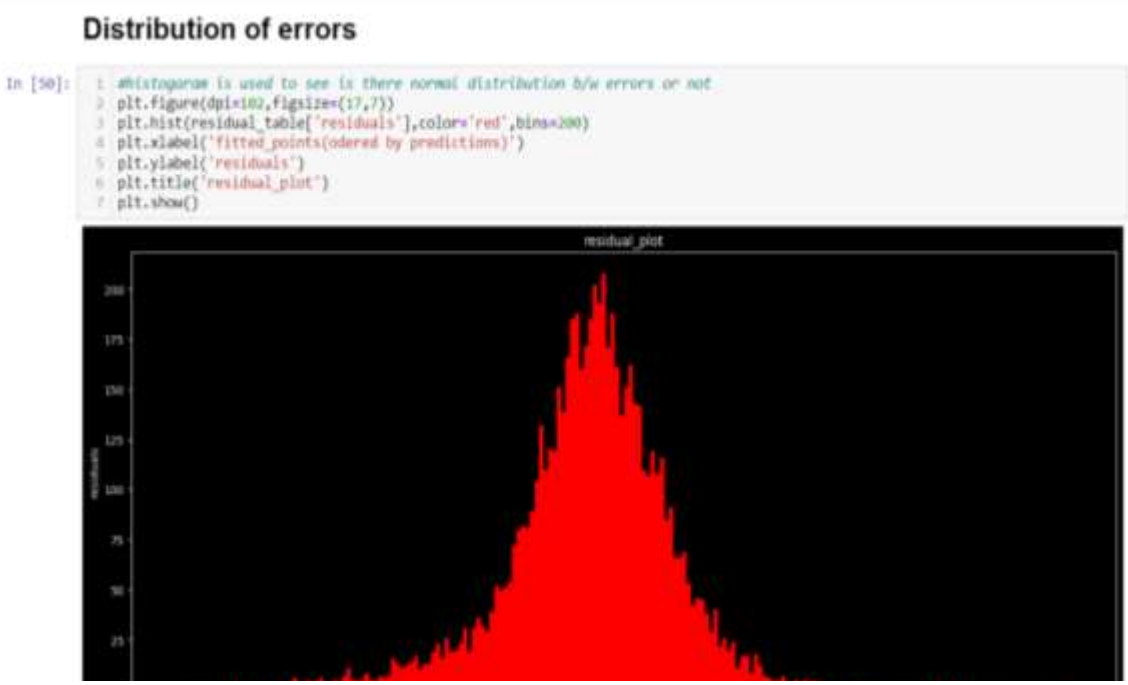

Fig4: Error distribution is Bell-shaped curve

VI. RESULT

The outcomes of a linear regression version for residence sale prediction ought to encompass the subsequent:

- Coefficient Estimates: The anticipated coefficients of the independent variables inside the model, which show the connection between each independent variable and the sale price.
- Goodness-of-healthy Measures: Statistical measures that compare how properly the version suits the information, including R-squared, adjusted R-squared, and the F-statistic.
- Predictions: Predictions for the sale fee of latest homes, based on the enter values for the unbiased variables.
- Residual analysis: A plot of the residuals (the difference among the located sale fee and the predicted sale charge) to make sure that the residuals are randomly disbursed and meet the assumptions of linear regression.
- version overall performance: A contrast of the performance of the linear regression version

with different models or benchmarks, the usage of measures inclusive of mean absolute blunders, suggest squared errors, and root suggest squared errors.

- Variable importance: An assessment of the significance of every impartial variable within the model, based totally on the value and significance of their coefficients.
- Limitations and Assumptions: A discussion of the limitations and assumptions of the linear regression model, which include linearity, independence, homoscedasticity, and normality, and how they affect the outcomes.

VII. CONCLUSION

A precis of the consequences of the linear regression version, consisting of the coefficient estimates, goodness-of-suit measures, and predictions for brand new records. An assessment of the overall performance of the linear regression model in predicting housing sale charges, based on measures along with suggest absolute errors, suggest squared error, and root suggest squared mistakes. A dialogue of the importance of every impartial variable in the version and how they impact the sale fee. A dialogue of the results of the effects for practitioners within the subject of housing, which includes real property retailers, developers, and traders. A concise typical conclusion on the effectiveness of linear regression in predicting housing sale charges, based totally at the consequences and overall performance of the model.

The reason of this observe is to deepen the expertise in regression techniques ingadget learning. device studying in computer science attempts to solveproblems algorithmically rather than basically mathematically. This takes a look at attempts toanalyses the correlation among variables to determine the most critical elementsthat have an effect on house prices. helps to discover the quality prediction for the residence the use of givendataset and mastering through it. And web App presents responsive consumer interface.

REFERENCES:

[1]. Martín Abadi, AshishAgarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian
[2]. (2015). https://www.tensorflow.org/ Software available from tensorflow.org.
[3]. Linear regression form java t point
[4]. features. arXiv:1609.08399[cs.CV] (2016).
[5]. S Arietta, Afros, R Ramamooorthi, and M Agrawala. 2014. City Forensics:
[6]. Using Visual Elements to Predict Non-
Visual City Attributes. IEEE Transactions
[7]. on Visualization and Computer Graphics (2014).
[8]. Herbert Bay, TinneTuytelaars, and Luc Van Gool. 2006. Surf: Speeded up robust.
[9]. Linear Regression model image referred from javaTpoint website.
[10]. Positive and Negative linear regression image line referred from javaTpoint website