IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Design And Implementation Of A House Price Prediction Using Machine Learning.

Mrs.C.Navamani M..E.,

Assistant Professor, Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India.

L. Dhashinesh

UG Scholar, Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India.

A. Midhunesh

UG Scholar, Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India.

N. Ranjith

UG Scholar, Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India.

D . Santhosh

UG Scholar, Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India.

ABSTRACT : Since previous years, machine learning has become increasingly important in major fields, normal speech commands, and product recommendations. Rather than this, it offers safer automotive systems and superior customer services. Given that machine learning (ML) is a popular technology across nearly all industries, we are attempting to include ML into our project. The real estate market is one of the most competitive these days in terms of price and volatility. People are searching to get a new house by analyzing market techniques and sticking to their finances. However, the primary drawback of the existing approach is that it determines a home's valuation without making the essential predictions about future market trends, which drives up prices. Thus, the primary goal is to estimate home prices accurately and profitably. Predicting an accurate house price requires taking into account a number of variables, including the

customer's preferences and budget, in order to provide effective house pricing. Thus, we are developing a model to forecast housing costs. by utilizing linear regression and other machine learning algorithms.

KEYWORDS: Linear regression, price Detection System, Machine Learning

I. INTRODUCTION

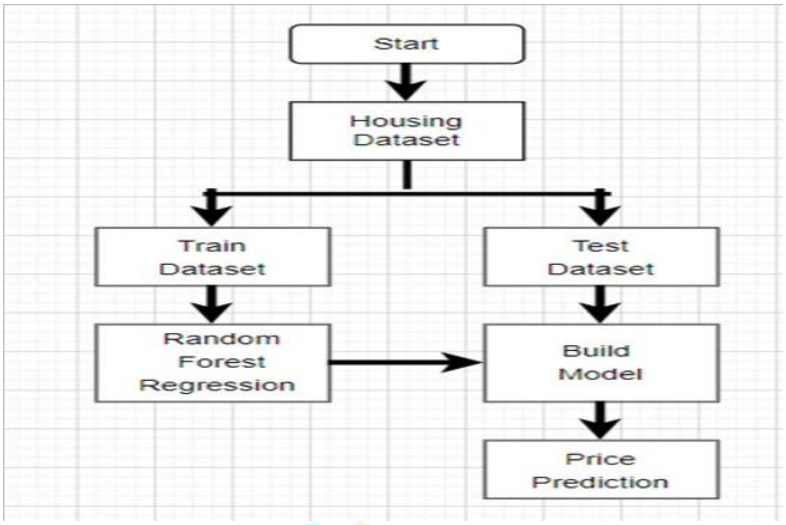
Within the field of economic analysis, the real estate market is a key player, providing profitable investment opportunities as well as a window into the health of the economy. In addition to reflecting the state of the economy, the fluctuating nature of housing values offers a wide range of options to investors, sellers, purchasers, and legislators. But the nuances of precisely predicting these prices are sometimes beyond the reach of conventional approaches, which mainly depend on past performance and professional opinion, frequently overlooking the complex interactions between different components. Machine learning algorithms have ushered in a new age in the forecast of property prices by utilizing large datasets and automated learning methods. A thorough investigation was conducted recently using a variety of machine learning methods, such as Linear regression . Decision Tree, K-Means, and Random Forest Regression analyses. Eighty percent of the carefully selected data in the dataset was put aside for rigorous training, while the remaining twenty percent was set aside for careful testing.

II. RELATED WORK

A number of research works that explore the nexus between data science and real estate have looked into the prediction of home values through machine learning methods. One prominent line of inquiry is modeling the intricate correlations between different variables and home prices by using ensemble techniques like random forests, decision trees, and regression algorithms like linear regression. To train and assess these models, researchers frequently make use of big datasets that include details on location, square footage, the number of bedrooms and bathrooms, amenities, neighborhood demographics, and economic factors. A key role is played by feature engineering, which uses methods like dimensionality reduction and the creation of new variables through interactions or transformations. Furthermore, some research looks into integrating cutting-edge methods like deep learning, which may identify complex patterns in the data but frequently need for bigger computational resources and datasets. Metrics like as mean absolute error, mean squared error, or coefficient of determination (R-squared) are commonly used to evaluate model performance, while cross-validation techniques are used to evaluate generalizability. Even with these developments, problems remain, such as problems with data quality, interpretability of models, and mitigating biases present in historical data. Machine learning-based house price prediction models have the potential to improve in accuracy and practicality through ongoing refinement and investigation of innovative techniques.

III. PROPOSED SYSTEM

3.1 FLOWCHART



We use machine learning methods such as Leaner Regression, Decision Tree, k-Means, and Random Forest to estimate housing prices in this suggested system. We suggested the "House Price Prediction Using Machine Learning" system, in which we make several feature-based house price predictions. We are able to train the model in this suggested manner using a variety of features. Eighty percent of the previously collected data is used for training, and the remaining twenty percent is used for testing. Hare, a ".csv" file contains the raw data. To tackle these issues, we primarily used two machine learning libraries. Pandas was the first, while Numpy is the second. In addition to being used to load ".csv" files into Jupiter Notebook, pandas and both to alter and to clean the data. Another was sklearner, which was employed for actual analysis and offers a number of built-in functions that aid in issue solving. Another library that is nothing more than numpy was utilized. Numpy was utilized in the train-test splitting process.

3.2 NUMPY

A key library for Python scientific computing is called NumPy (Numerical Python). It is an essential tool for data analysis and machine learning jobs since it offers strong tools for working with multi-dimensional arrays and matrices. Here's a closer look at what NumPy can do:

Essential Features:

Arrays with N dimensions: Arrays, which are effective data structures that may hold elements of the same data type in various dimensions, are introduced by NumPy. You can now represent intricate data structures like matrices and tensors thanks to this.

Array Operations: Compared to conventional Python loops, it enables a vast array of mathematical functions that operate on full arrays simultaneously, greatly enhancing efficiency. These functions may

perform a wide range of tasks, including random number generation, linear algebra, and element-wise calculations.

Broadcasting is a potent technique that makes working with arrays of various forms easier. In order to reduce the need for human loops, NumPy automatically expands or replicates arrays to execute element-wise calculations.

NumPy arrays are frequently used in data preprocessing, which involves reshaping, cleaning, and modifying data for machine learning models.

Linear Algebra: NumPy offers efficient functions for matrix operations like as multiplications, inversions, and decompositions, which are crucial for a variety of machine learning techniques.

Array Manipulation: NumPy arrays are useful for feature engineering and model computations because they can handle complex mathematical operations on huge datasets.

3.3 PANDAS

Another significant Python module made especially for data analysis and manipulation is called Pandas. Building upon NumPy, it adds more advanced data structures and tools while retaining NumPy's effective array functionality. Here's how NumPy and Pandas work together to enhance data science tasks.

Fundamental Data Structures:

Data Loading: Pandas simplifies data import by providing functions to read data from a variety of file formats, including databases, CSV, Excel, and JSON.

Data cleaning: It gives you tools to deal with duplication, deal with missing values, and clean up inconsistent data.

Data Transformation: Filtering, sorting, grouping, and feature engineering chores are all made efficient with Pandas.

Data Analysis: It has its own statistical functions for data exploration and analysis, and it interfaces with other libraries such as NumPy for numerical computations.

Pandas is an excellent tool for cleaning and preprocessing data before putting it into machine learning models. It can handle missing values, outliers, and inconsistent data. It turns current data into new features that are appropriate for training models, hence simplifying the process of creating features. Pandas makes it easy to explore and analyze data to find patterns and relationships between variables, which can help with the selection and understanding of models.

3.4 EXPERIMENTAL EVALUATION

To determine how well machine learning models anticipate home prices, experimental evaluation is essential. Here's a summary of the crucial actions:

Gathering and Preparing Data:

• Gathering of Data: Obtain a dataset of property prices that includes attributes such as area, number of bedrooms, location, amenities, and the final selling price.

• Preprocessing: Organize the data by addressing missing values, encoding categorical categories, and maybe scaling numerical features to make sure they are equivalent in magnitude.

Model Training and Model Selection:

• Model Selection: For regression problems, select suitable machine learning techniques, such as Gradient Boosting, XGBoost, Random Forest Regression, or Linear Regression.

Model Evaluation:

• Performance measures: Use measures such as these to assess the model's effectiveness on the test data that hasn't been seen.

A measure of the average squared difference between expected and actual prices is the mean squared error, or MSE. Performance is improved when the MSE is lower.A measure of the average absolute difference between expected and actual prices is the mean absolute error, or MAE. A lower MAE is a sign of superior performance.R-squared: Shows how much of the variation in the real prices can be accounted for by the model. A better match is shown by a number that is closer to 1. Comparing: Determine which model has the most prediction potential by comparing the performance of various models using the selected metrics.4. Keep an eye out for overfitting, a situation in which a model performs well on training data but badly on unobserved data. This can be lessened with the use of strategies like regularization and hyperparameter optimization.

• Feature Importance: Examine the model's analysis to see which features have the biggest impact on the prediction, offering insights into the variables affecting home prices.

IV. CONCLUSION AND FUTURE WORK

A useful resource for predicting home prices is machine learning. Models that assess the worth of new assets can be trained by examining past data and finding critical characteristics that impact pricing. Many parties involved in the real estate industry may profit from this. Customers are able to make wise judgments by learning about fair market value.Decide on reasonable asking prices to perhaps speed up sales.Determine undervalued properties by analyzing possible investments.In extremely volatile markets where prices fluctuate quickly, predictions might not come true.Training data completeness and quality have a major impact on the model's performance. Explaining the predictions of certain models (like deep learning) might be tough since they are intricate and hard to comprehend. Machine learning for predicting housing prices is a rapidly developing field. The following areas show promise for further research. Utilizing Outside Data is

Predictive accuracy can be increased by incorporating data on local infrastructure projects, demography, and the economy. Building models that can evaluate current market trends and modify forecasts in response is known as real-time market analysis. Develop models that are easier to grasp so that users may gain confidence in the system and comprehend the logic underlying the predictions. This is the main goal of explainable AI (XAI).

V. REFERENCES:

[1] Rohini Nair - "Housing Price Prediction Using Machine Learning and Neural Networks" 2018, IEEE.

[2] House Price Prediction Using Regression Techniques: A Comparative Study" 2019 in (ICSSS),IEEE.

[3] Sifei Lu, Zengxiang Li, Zheng Qin , Xulei Yang , Rick Siow Mong Goh - "A hybrid regression technique for house prices prediction" 2017 – IEEE.

[4] Bharatiya, Dinesh, et al. "House price prediction using linear regression." Electronics, Communication, and Aerospace Technology(ICECA),2017 International conference of. Vol. 2. IEEE, 2017.

[5] "Machine Learning in Action" by Peter Harrington-Machine-Learning (ICECA) , VOL 1 - IEEE , 2019.

[6] "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy.

[7] A Machine Learning Approach to Housing Price Prediction by Hongyu Wu, Xiangxiang Kong, and Wenfei Wu.

[8] Machine Learning Methods Applied on House Price Prediction: A Survey By -Ke Li, Peng Cheng, et al.

[9] Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien Géron.