e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707115 |

Real Estate Price Prediction using Machine Learning

Ms. Barnali Chakraborthy, Bikash Mishra, Dr. Pavan.G P

Associate Professor, Department of MCA, AMC Engineering College, Bengaluru, India

4th Semester MCA, Department of MCA, AMC Engineering College, Bengaluru, India

Department of ISE, AMC Engineering College, Bengaluru, India

ABSTRACT: The abstract of the real estate price prediction project for properties in Bengaluru involves using machine learning algorithms to develop a predictive model that can estimate the prices of properties in the city based on various features such as location, size, amenities, and so on. The dataset used for this project contains information on thousands of properties in different parts of Bengaluru, including their sale prices, size, location, number of bedrooms, bathrooms, and other key features. To build the predictive model, several machine learning algorithms such as linear regression, decision tree, random forest, and XGBoost are used to train the model on the given dataset. The performance of each model is evaluated using metrics such as mean absolute error, mean squared error, and R squared score, and the best performing model is selected as the final predictive model. Once the model is trained and tested, it can be used to predict the prices of new properties in Bengaluru based on their features. The model can also be used to identify the most important features that affect the price of properties in the city, which can be useful for real estate agents, property developers, and investors looking to buy or sell properties in Bengaluru.

KEYWORDS: Machine Learning, Location, Amenities, Regression, Absolute.

I. INTRODUCTION

Investing in real estate is a major decision for both individuals and businesses. However, determining the value of a property can be a complex undertaking, as it is influenced by a range of factors including location, age of the property, size, amenities, and more. Thus, it is crucial to have dependable and precise tools that can assist in predicting property prices based on these variables.

Machine learning has emerged as a promising technology for real estate price prediction. Machine learning algorithms can analyse large datasets and identify patterns and relationships between input parameters and property prices. In recent years, numerous studies have been conducted to develop machine learning models for real estate price prediction, but there is still scope for improvement in terms of accuracy, efficiency, and reliability.

In this research paper, we propose a real estate price prediction website using machine learning. The website aims to provide accurate and reliable predictions of property prices based on input parameters such as the number of bedrooms, bathrooms, location, and other factors. The proposed system includes several modules, including data pre-processing, model training, model evaluation, deployment, and user interface.

The data pre-processing module involves cleaning and transforming the raw data to make it suitable for use in the model. The model training module uses supervised learning and linear regression to build the model and gradient descent to optimize the model's parameters. The model evaluation module evaluates the performance of the trained model using various techniques such as crossvalidation and learning curves. The deployment module deploys the trained model using Flask, and the user interface module creates an attractive and user-friendly interface using HTML, CSS, and JavaScript.

II. LITERATURE SURVEY / EXISTING SYSTEM

The purpose of this system is to determine the price of a house by looking at the various features which are given as input by the user. These features are given to the ML model and based on how these features affect the label it gives out a prediction. : [1] Regression Techniques: Many studies employ various regression models such as linear regression, support vector regression (SVR), decision trees, and random forests. These models are used to predict real estate prices

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707115 |

based on features such as location, size, number of rooms, amenities, etc. Researchers often compare the performance of these models to find the most accurate one for predicting prices. [2] Feature Selection: Feature selection plays a crucial role in improving prediction accuracy. Researchers explore different sets of features to determine which ones have the most significant impact on price prediction. Techniques like principal component analysis (PCA) and feature importance analysis in tree-based models are commonly used.[3] Geospatial Analysis: Some studies incorporate geospatial data such as proximity to schools, transportation hubs, crime rates, and neighborhood characteristics. Geospatial features can significantly enhance the predictive power of models by capturing location-specific nuances.[4] Time-Series Analysis: Real estate prices often exhibit temporal trends. Time-series analysis techniques, including autoregressive integrated moving average (ARIMA) and exponential smoothing methods, are used to model and predict price movements over time. [5] Deep Learning Approaches: Recent advancements in deep learning have also been applied to real estate price prediction. Techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can capture complex patterns in real estate data, potentially outperforming traditional machine learning models in certain scenarios.

[6] Data Preprocessing and Cleaning: High-quality data preprocessing is essential for accurate predictions. Techniques like outlier detection, missing value imputation, and normalization are applied to ensure the reliability of input data. [7] Ensemble Methods: Ensemble methods like bagging, boosting (e.g., gradient boosting machines), and stacking are frequently used to combine multiple models and improve prediction accuracy.[8] Evaluation Metrics: Common evaluation metrics include mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R-squared). These metrics help researchers quantify the performance of their models and compare them against benchmarks. [9] Case Studies and Applications: Literature often includes case studies where these models are applied to real-world datasets from different cities or regions. This application demonstrates the effectiveness and generalizability of the proposed methods. [10] Challenges and Limitations: Studies also discuss challenges such as data heterogeneity, data scarcity (especially in less developed regions), model interpretability, and overfitting. Addressing these challenges is crucial for deploying robust real estate price prediction systems.

To conduct a comprehensive literature survey, you would typically review research papers, conference proceedings, and academic journals focusing on real estate economics, machine learning, and data science. This approach helps in understanding the current state-of-the-art methods and identifying opportunities for further research and improvement.

III. PROPOSED METHODOLOGY AND DISCUSSION

The purpose of this system is to determine the price of a house by looking at the various features which are given as input by the user. These features are given to the ML model and based on how these features affect the label it gives out a prediction. This will be done by first searching for an appropriate dataset that suits the needs of the developer as well as the user. Furthermore, after finalizing the dataset, the dataset will go through the process known as data cleaning where all the data which is not needed will be eliminated and the raw data will be turned into a .csv file. Moreover, the data will go through data preprocessing where missing data will be handled and if needed label encoding will be done.

1. Data Collection and Preprocessing:
- Data Sources: Gather real estate data from reliable sources such as property listings, real estate websites, government databases (if available), and third-party data providers.
- Data Cleaning: Clean the data to handle missing values, outliers, and inconsistencies. Ensure data quality through rigorous preprocessing steps including normalization, feature scaling, and encoding categorical variables.

2. Feature Selection and Engineering:
- Feature Selection: Use techniques like correlation analysis, feature importance from tree-based models, or domain knowledge to select the most relevant features (e.g., location, property size, number of bedrooms/bathrooms, amenities, neighborhood characteristics).
- Feature Engineering: Create new features that may enhance predictive power, such as proximity to schools, hospitals, public transport, crime rates, and economic indicators.

3. Model Selection and Training:
- Model Types: Evaluate and select appropriate regression models based on the nature of the problem (e.g., linear regression, support vector regression, decision trees, random forests, gradient boosting machines).
- Hyperparameter Tuning: Optimize model performance through techniques like grid search or Bayesian optimization to find the best hyperparameters.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707115 |

4. Validation and Evaluation:
- Cross-Validation: Use techniques such as k-fold cross-validation to assess model generalizability and reduce overfitting.
- Evaluation Metrics: Measure model performance using metrics like mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and R-squared to quantify prediction accuracy.

5. Deployment and Monitoring:
- Deployment Strategy: Implement the trained model into a production environment, ensuring scalability and reliability.
- Monitoring: Establish mechanisms to monitor model performance over time and update the model periodically as new data becomes available.

Discussion

1. Choice of Models and Features: Justify the selection of specific machine learning models based on their suitability for real estate price prediction tasks. Discuss why certain features (e.g., location, amenities) are chosen and their expected impact on prediction accuracy.
2. Data Challenges and Solutions: Address challenges encountered during data collection and preprocessing, such as data heterogeneity and missing values. Explain how these challenges were mitigated to ensure data quality and model reliability.
3. Model Performance and Comparison: Present results from model evaluation and compare performance metrics across different models considered. Discuss any trade-offs between model complexity and predictive accuracy.
4. Practical Implications and Applications: Highlight potential applications of the proposed methodology in real-world scenarios, such as assisting real estate agents in pricing properties, supporting investors in making informed decisions, or aiding policymakers in understanding housing market trends.
5. Limitations and Future Directions: Acknowledge limitations of the proposed methodology, such as data availability constraints or model interpretability issues. Propose avenues for future research, such as incorporating more advanced machine learning techniques (e.g., deep learning) or integrating additional data sources for enhanced prediction capabilities.
6. Ethical Considerations: Discuss ethical considerations related to using machine learning for real estate price prediction, including potential biases in data, transparency of model predictions, and implications for fairness in housing markets.

By thoroughly discussing these aspects, you can provide a comprehensive overview of your proposed methodology for real estate price prediction using machine learning, demonstrating its relevance, effectiveness, and potential impact in the field.

IV. EXPERIMENTAL RESULTS

- Represent the method of working function

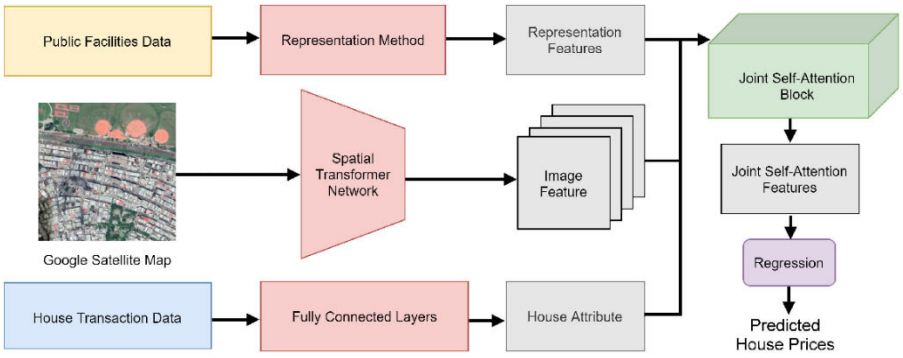

Fig.4.1

.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707115 |

- Evaluation of price of house.

Fig.4.2

- Condition work in software.

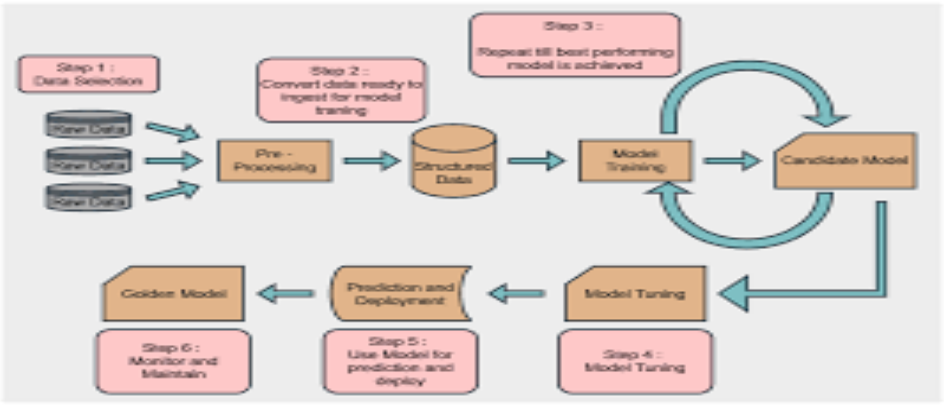

Fig.4.3

V. CONCLUSIONS

Thus the machine learning model to predict the house price based on given dataset is executed successfully using xg regressor (a upgraded/slighted boosted form of regular linear regression, this gives lesser error). This model further helps people understand whether this place is more suited for them based on heat map correlation. It also helps people looking to sell a house at best time for greater profit. Any house price in any location can be predicted with minimum error by giving appropriate dataset.

REFERENCES

1. Real Estate Price Prediction with Regression and Classification, CS 229 Autumn2016 Project Final Report
2. Gongzhu Hu, Jinping Wang, and Wenying Feng Multivariate Regression Modellingfor Home Value Estimates with Evaluation using Maximum Information Coefficient
3. Byeonghwa Park , Jae Kwon Bae (2015). Using machine learning algorithms for housing price prediction , Volume 42, Pages 2928-2934 [4] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, 2015. Introduction to Linear Regression Analysis.
4. Iain Pardoe, 2008, Modelling Home Prices Using Realtor Data • Aaron Ng, 2015, Machine Learning for a London Housing Price Prediction Mobile Application
5. Wang, X., Wen, J., Zhang, Y.Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. Optik-International Journal for Light and Electron Optics, 125(3), 14391443

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com