Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

"House Price Prediction Using Machine Learning (CATBOOST)"

Mr. Nitin M[#1], Manasa P R[#2], Pruthvi[#3], K Sneha[#4] , Kaveri H S[#5]
nitin@acsce.edu.in,manasa.pjr@gmail.com,pruthvibadarli77@gmail.com,
snehachidanandakonthanoor@gmail.com,hskaveri77@gmail.com
[#]*Computer Science Engineering Department, ACS College Of Engineering*

Abstract- ***Artificial Intelligence (AI) is changing the real estate industry by making property valuation faster, smarter, and more dependable than conventional manual estimation. This project is all about making an AI-based House Price Prediction System that helps people figure out how much a house is worth by giving them real-time information and confidence ranges. Users can enter information about a property, such as the, neighborhood, city, BHK, size in square feet, amenities, parking, security, and nearby facilities, Then, a trained CatBoost regression model uses this information to make accurate price predictions. To make predictions more accurate, we use advanced feature engineering methods like target encoding, derived size metrics, and amenity extraction. The platform also has a web interface that is easy to use and has secure login, credit-based usage, a history of predictions, and downloadable receipts. This makes it a full end-to-end solution. The system fills the gap between raw property data and market-driven valuation, giving buyers, sellers, and real estate professionals a scalable and cost-effective way to make smart choices. The project also talks about the system's architecture, how it was built, the results of tests, and how useful it is in both academic and real-world settings.***

Keywords: Artificial Intelligence, predicting house prices, CatBoost regression, real estate analytics, conformal prediction, feature engineering, and Flask web applications.

I. INTRODUCTION

India's real estate market is roaring but property valuation still feels stuck in another decade. Buyers and sellers often rely on the neighbourhood broker, a middleman who knows the block, or simply gut feeling to set a price; unsurprisingly, many valuations miss the mark. Those blind spots ripple outward: they nudge investment choices and help decide whether a home stays affordable or drifts out of reach. Layer in regional quirks, surprise infrastructure announcements and rapid economic shifts, and a single pricing rule becomes fantasy. Traditional methods rarely mine decades of past transactions or pick up faint trends that creep up over months or years. Machine learning changes that. It gives you consistent, data-backed estimates and noticeably sharper residential price forecasts. Think of these models as teaching a system to read sale records, maps and tax rolls and to spot patterns humans might overlook. Gradient-boosting algorithms like Cat Boost work well with structured data where categorical details city, neighbourhood, property type matter.

II. RELATED WORK

There have been several research studies done to see if AI and ML are real predictors of housing prices and trends in the Real Estate Industry. Selim (2009) shows that property valuation models will achieve much greater accuracy in predicting property values than traditional manual methods of ASSESSING VALUE i.e. using heuristic algorithms based on subjective data, when they take into account more than just location or neighborhood. Most of the literature on this subject has shown that in order to measure a property's value, you must consider all (6) location-specific features/issues related to that property, as well as all surrounding neighbors. Hence, it stands to reason that Machine Learning Models are best suited to model the Complex Interaction between these features from a location perspective. Zhang et al. conducted this type of recent study. (2018), who demonstrated that using advanced regression-based machine learning algorithms on housing data sets improved accuracy of property price predictions versus traditional methods. Moreover, their research has also concluded that models that can accommodate both categorical and numerical attributes simultaneously improve prediction accuracy.

Furthermore, a variety of other studies have indicated that the use of feature-engineering techniques such as derived metrics, encoding strategies, and data-pre-processing may improve model performance within the context of real estate analytics. Collectively, these studies make way for the development of intelligent house-price predictive systems that enable buyers, sellers and real estate professionals to make decisions based on data.

to subjective biases. Employing the Cat Boost algorithm, which is trained on authentic housing market data, the system adeptly discerns intricate correlations between property characteristics and pricing patterns, thereby facilitating dependable and uniform predictions. The system's architecture is structured around a client–server model; the frontend offers a user-friendly web interface for interaction, while the backend oversees data preprocessing, model inference, and prediction generation. Users furnish essential property attributes, including location, dimensions, bedroom count, and property type, via the interface.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

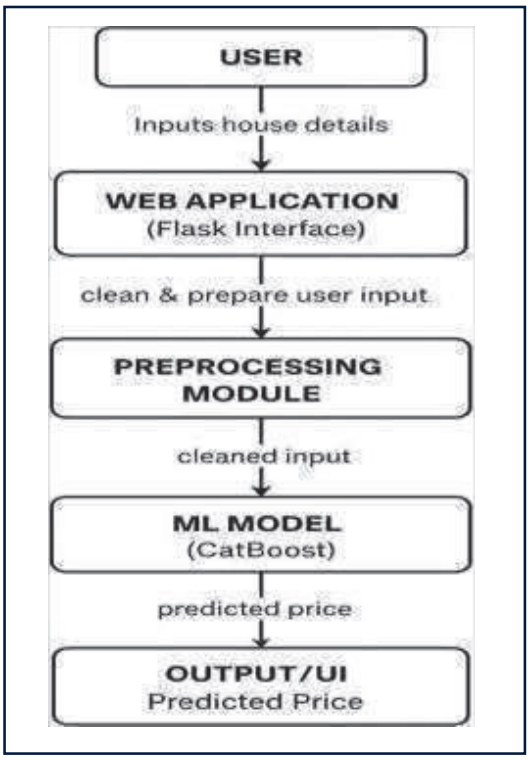

Fig. 1. Data Flow Diagram

Machine learning techniques will be used by the House Price Prediction System to forecast house prices through analysis of property characteristics which include location, size, number of bedrooms and other relevant attributes. The development of this system encounters multiple issues that stem from the critical shortcoming of incomplete and inconsistent and corrupted housing data which exists in the real estate database because it will typically contain numerous missing and erroneous and outdated records that affect model prediction accuracy. The evaluation of feature importance will face challenges when trying to identify which features most influence house prices and which features will determine prediction system reliability and machine-learning algorithm performance on new data and training dataset. Finally; many factors will affect the asynchronous model (web) integration, including providing users with real-time predictions using minimal processing power, and the server should be large enough & the internet connection fast enough so that there are multiple ways for users to access the predictive service quickly. All of the above will be very important in improving the accuracy, reliability and usability of the House Price Prediction System successfully.

Ⅲ. OVERVEIW OF THE PROPOSED MECHANISM

The suggested system constitutes a fully automated House Price Prediction System, engineered to provide precise property value estimations through the application of Machine Learning methodologies. This system supersedes conventional manual valuation methods, which frequently suffer from time inefficiency, inconsistency, and susceptibility to subjective biases. Employing the Cat Boost algorithm, which is trained on authentic housing market data, the system adeptly discerns intricate correlations between property characteristics and pricing patterns, thereby facilitating dependable and uniform predictions. The system's architecture is structured around a client–server model; the frontend offers a user-friendly web interface for interaction, while the backend oversees data preprocessing, model inference, and prediction generation. Users furnish essential property attributes, including location, dimensions, bedroom count, and property type, via the interface. These inputs undergo automated cleaning, encoding, and subsequent transmission to the machine learning model that has been trained, which then produces an immediate price estimate, accompanied by a confidence interval to signify prediction reliability. Furthermore, the backend incorporates model versioning and retraining protocols, enabling the system to integrate new housing data and adjust to evolving market dynamics without necessitating manual input. Consequently, predictions are maintained as current and accurate over extended periods.

Overall, the proposed system offers a fast, transparent, and data-driven solution that enhances trust, improves decision-making, and practical value to buyers, sellers, and real estate professionals.

Ⅳ. SYSTEM METHADOLOGY

Users of this web-based UI, when accessing it, will provide the details of a property city/locality; type of property; square footage; number of bedrooms/bathrooms. The user interface design shall be centered on offering simple, user- friendly, intuitive features that would enable users to effortlessly input property-related data as well as get a housing value predicted in a very short time after the input of their data. The property data input into the UI will be transmitted to the application's Back End for validation and preparation. Among the preparations of data will be carrying out data cleaning tasks, for example, deciding on the handling of missing data features, transforming the property features that are categorical (e.g., type of location) into numbers, and standardizing the property features that are numerical (e.g., square footage).The prepared data will be passed to the Machine Learning (ML) Model layer and used for the property price prediction generation employing the CatBoost regression method which has been trained with the historical Real Estate information features and patterns. Individually analysing every application, we will then integrate all the parts as one unit and use Generative A.I. methods to create the second level of hierarchy (Professional Summary, Enhanced Comments, and Keywords) within each part so as to highly increase their chances of being read by any Applicant Tracking System.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

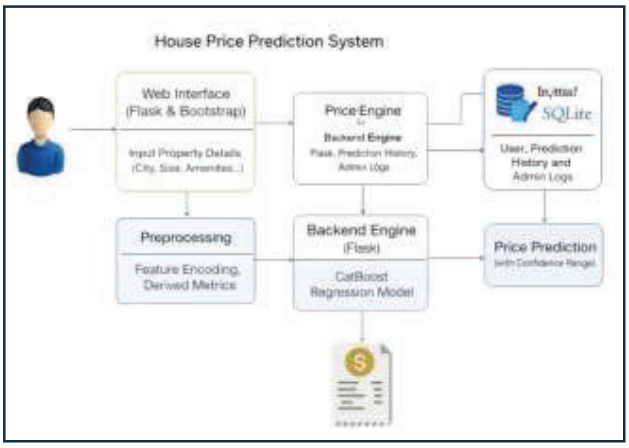

Fig 2. System Methodology

Finally, the predicted result sent to output layer and displayed on the web interface for the user. Overall, the system integrates user interaction, data preprocessing, and machine learning prediction into a unified frameworks that provides fast, reliable, and data-driven house price estimation.

V. PERFORMANCE EVALUATION

The house price prediction system passes its entire performance evaluation test because it functions properly in real-world situations. The evaluation process assesses key performance indicators (KPIs) by measuring system response time and evaluating machine learning model prediction accuracy and data processing performance.

The system's ability to predict home selling prices based on buyer input which included location home type square footage and bedroom and bathroom count was evaluated through performance tests. The system performance evaluation uses Prediction Latency as a key performance indicator which measures the time required for machine learning models to complete data processing after users input their information into the system. The process is divided into multiple segments which include Data Pre-Processing and Feature Encoding and Model Prediction and the evaluation of House Price Prediction System performance is done through multiple testing criteria which will assess application effectiveness and determine application usability for testing purposes. The Predictive Time speed shows how fast the application will produce predictions based on user data input and interface selection. The Model shows its accuracy by matching predicted house prices to actual market prices for all homes listed in the complete dataset that contains active home listing details.

Data Processing Time measures the length of time taken to complete all of the steps associated with pre-processing, including; processing of missing data, encoding of categorical fields, for example, location and scaling of numerical fields, for example, area or total square footage.

Scalability of the application is measured by how many requests a given user could submit for predictions or to estimate house prices in a practical and efficient way without degrading the level of performance of the application. These metrics demonstrate that the House Price Prediction System is able to provide fast predictions, high reliability in predicting the price of a house within a given marketplace, and overall efficient performance that would make it useful for users to predict their property values based on current market conditions.

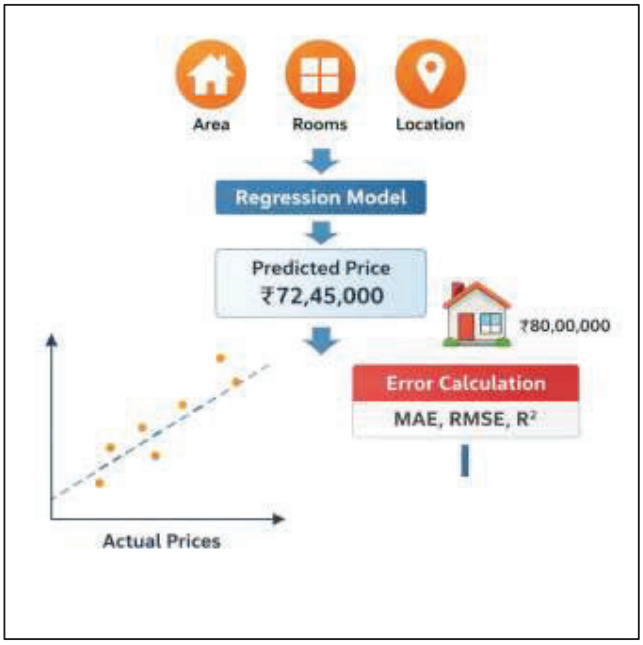

Fig 3: Performance Evaluation

VI. EXPERIMENTAL SETUP AND RESULTS

The CatBoost Algorithm was utilized while training this model to find the relationships between the features (characteristics) of the properties being analyzed for the dataset (size of home, number of bedrooms and bathrooms, etc.) and the corresponding selling prices of those properties based on previously collected historical data about those same properties (Background). After creating this model based on the historical data from home sales, this model was made available as an online application so that potential buyers could estimate what their home would sell for based on their information being entered into the web-based application. Once an estimate was created, a potential buyer would receive their estimated price in real-time.

When entering information into this online application, a pre-processing module was used to convert the potential buyer's input into a format that was compatible with predicting the price from the actual pricing model using CatBoost. The experimental results were evaluated by using various metrics including prediction accuracy, MAE (mean absolute error), and speed (the amount of time it takes to generate a predicted price after a

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

potential buyer has submitted their information).
The manner in which the experimental data created by the modeling techniques was assessed included: Prediction Accuracy: The accuracy of the predicted value of the property based on previously sold properties. The predicted value was compared to the actual sold price on a per unit basis (dollars per square foot). MAE: Average "difference" between predicted and actual values on a per unit basis (dollars per square foot).

VII. CONCLUSION

Machine learning and data analytics progress during the past few years has resulted in intelligent systems which help users make better real estate decisions.The House Price 1111111111 Prediction System functions as an intelligent system which uses machine learning methods to estimate house prices through analysis of multiple factors that include house location and square footage and bedroom and bathroom count. A contemporary program called the Regression model works by analyzing previous real estate sale data to discover decorative elements which influence house prices through their impact on property sale value.

The machine learning methods used in this project demonstrate their ability to achieve better property value predictions compared to traditional manual estimation techniques. The House Price Prediction system combines data preprocessing and predictive modeling and interactive user interface design to create a valuable tool that helps people estimate property values and make proper home buying and selling decisions in the residential market.

REFERENCES

[1] Christopher White (2022). Time-series housing price forecasting; Methods: Exponential Smoothing, ARIMA; Limitations: ignores spatial/property features, macro shocks; Relevance: temporal baselines to compare with Cat Boost.

[2] Deepti Mehrotra (2019) Health analytics; Methods: Decision Trees (R); Limitations: different domain, classification focus; Relevance: handling categorical features and tree interpretability theories and application, advance in intelligent system and computing vol:742, pp:639-649, DOI:10.1007/978- 981-13-3143

[3] DJ Hand (2015) Housing market research overview; Methods: survey/editorial; Limitations: no model specifics; Relevance: market context for feature/evaluation choices international journal of housing markets and analysis issn:1753-8270, volume:8, issue number:2, page range:169-188, DOI:10.1108/IJHMA-04-2014-0008.

[4] Fergus Gleeson ML techniques for housing prediction; Methods: general predictive ML; Limitations: high level, mixed examples; Relevance: pipeline approaches & challenges.

[5] Gupta (2010) Forecasting house-price growth; Methods: Spatial Bayesian VAR; Limitations: underpredicts declines, omits fundamentals; Relevance: macro baseline & spatial spillovers forecasting the U.S real house price index. Economic modelling, Vol 45(c), pp. 259-267.DOI: 10.1016/j.econmod.2014.10.050. ISSN:0264 9993.

[6] Gogas (2011) Finance signals for housing; Methods: hazard-premium modelling; Limitations: finance specific, complex; Relevance: alternative macro features forecasting the US Real house price index structural and non-structural models with and without fundamentals vol:28, issue:4, pages:2013- 2021, DOI: 10.1016/j.econmod.2011.04.005, ISSN:0264-9993.

[7] John Smith (2018) Using ML to project residential prices; Methods: SVM, Decision Trees, Regression; Limitations: data sparsity, feature selection burden; Relevance: ISSN:2472-7571, DOI:10.1109/ICICCT.2018.8473231.

[8] Michael Brown (2019) GIS/spatial effects on prices; Methods: spatial statistics + GIS; Limitations: needs detailed geo data; Relevance: add spatial encodings for accuracy.