Home Construction Cost Estimation Using ML

SUCHETHA N V[1], ADITHYA[2], ASHIK S[3], DHANUSH[4], THRINESH REDDY S GULEDAGUDDA[5]
[1]*Asst. Professor, Department of CSE, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India*
[2, 3, 4 , 5]*Student, Institute Department of CSE, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India*

Abstract- Machine learning has played a significant role in diverse fields, including speech recognition, product recommendations, and the medical industry. Its application has led to advancements in customer service and automobile safety, making it a widely adopted technology across various domains. With the constant fluctuations in housing prices, individuals are seeking to purchase new homes within their budget while analyzing market trends. However, the existing systems for predicting home costs suffer from a notable drawback as they do not consider future market trends, potentially leading to unexpected price increases. In light of this, our project aims to address this issue by developing a housing cost prediction model that eliminates losses and provides accurate estimations. To achieve this, we are utilizing several machine learning algorithms, namely Linear Regression, Gradient Boost Regression, and XGBoost Regression. By incorporating these algorithms, we aim to enable individuals to make informed property investments without the need for a broker. Through our research, we have determined that the Linear Regression algorithm yields the highest accuracy in predicting home costs.

Indexed Terms - accuracy, house cost, Housing prices, Machine Learning.

I. INTRODUCTION

Every single organization is working effectively in real estate business. We all know that building their dream home is a common passion for many youths. Demands for housing are also rapidly growing the years. Many people are investing their lifetime saved money towards the property and their dream house and some people are buying the house. Over long ago, all these procedures were done manually. But in the manual method 25% of error occurs and causes loss of money. Today ML and AI are the trending technologies as many industries are moved towards automation. There is a need to simply the process for a normal human being while providing the best results. This project proposes a system that predicts home cost using a machine learning.

Home cost estimation can be done by using multiple prediction models such as Ridge regression, Lasso Regression and more. This model is more beneficial to house builders, property makers, investors, and home buyers. This model provides a information about the valuation of home cost in the present market and also helps potential buyers to decide the characteristics of a house according to their wish and budget. This model consists of both predicting home costs and attributes together.

II. LITERATURE SURVEY

The authors [1] Thuraiya Mohd, and Suraya Masrom was findings, The Random Forest Regressor exhibited the highest accuracy, closely followed by the Decision Tree Regressor. Similarly, the Ridge and Linear Regression models yielded comparable results, with a marginal decrease observed in the Lasso Regression. Notably, there was no substantial disparity observed among different feature selection groups, regardless of their strength. This finding indicates that predicting selling prices solely based on buying prices is viable, without the need to consider additional features that may contribute to model over-fitting. This consistent pattern is also evident in the Root Square Mean Error (RMSE) values across all feature selections.

The author [2] Sayan Putatunda employed machine

learning techniques to predict prices by utilizing attribute variables. They conducted experiments on a real-world Construction dataset, employing advanced machine learning algorithms such as Random Forest, Gradient Boosting, and artificial neural networks. Through performance comparison, they concluded that the Random Forest method achieved the highest prediction accuracy among the tested methods.

The authors [3] B.Balakumar and P.Ravirajutilized machine learning algorithms for predicting home costs. They employed linear regression as the chosen algorithm and conducted their analysis on the Bengaluru house dataset. Their focus was on identifying the features that buyers are particularly interested in to determine the most suitable option for them.

The paper [4], as published by M Thamaraj and S P Malarvizhi in IEEE in 2020. This study introduces a novel approach to predict home costs in Chennai, India, employing machine learning techniques. The authors collected data from diverse sources such as property websites, real estate agents, and government websites to construct their model. Various feature selection methods were utilized to identify the most relevant factors for accurate home cost prediction. The study proposes a unique machine learning algorithm named Boosted Regression Tree (BRT) specifically designed for this purpose. To evaluate its performance, the authors compared BRT with popular algorithms like linear regression, decision tree, and random forest. The findings demonstrated that the BRT algorithm exhibited superior accuracy and robustness compared to the other algorithms. Additionally, the study conducted an extensive analysis of the significant features influencing home costs and their respective impact on the predicted outcomes.

The paper [5] "Home cost Prediction Using ML Model" by Ismail Ibrahim presented a study focusing on the prediction of home costs in Malaysia. The author employed two machine learning models, namely Random Forest and Extreme Gradient Boosting, to carry out the analysis. The performance of these models was compared in terms of accuracy and robustness, with the results indicating that the XGBoost model outperformed the Random Forest model. As a conclusion, the paper asserts that the XGBoost model represents a more effective approach for predicting home costs and provides valuable insights for the real estate industry. The study's findings have practical implications and can contribute to the development of pricing strategies for properties in Malaysia.

The paper [6] "Home cost Forecasting Using Machine Learning" by Kuvalekar, Alisha et al. This study presents a machine learning-based approach for forecasting home costs, utilizing a dataset of home costs and their corresponding features in India. The authors conducted a comparative analysis of three machine learning models, namely Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting (GB), focusing on their accuracy and robustness. The findings revealed that the GB model exhibited superior performance compared to the SVR and RF models. As a result, the paper concludes that the GB model serves as an effective approach for forecasting home costs, offering valuable insights for the real estate industry.

The paper [7] "Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data" In their study, Park and Bae investigate the prediction of housing prices in Fairfax County, Virginia, utilizing machine learning algorithms. The authors leverage a dataset containing various features such as location, number of bedrooms, and lot size to forecast housing prices. They conduct a comparative analysis of four machine learning models, including Decision Tree, Random Forest, Gradient Boosting, and Artificial Neural Network, with a focus on accuracy and robustness. The findings indicate that the Gradient Boosting model surpasses the other models in terms of performance. The paper concludes that machine learning algorithms offer precise predictions for housing prices and hold significant value as a tool for the real estate industry.

The paper [8] "Literature Review on Construction Value Prediction Using Machine Learning" In their work, Babu and Dr. Chandran present a comprehensive literature review focusing on the prediction of construction value using machine learning techniques. The authors emphasize the significance of accurately forecasting construction

value and explore the potential advantages of employing machine learning in this domain. They critically examine a range of studies that have employed machine learning models, such as regression models, artificial neural networks, and decision trees, to predict construction value. By doing so, the paper offers valuable insights into the current landscape of research on construction value prediction using machine learning, while also identifying the strengths and limitations of existing models. The study's findings hold the potential to guide future research endeavors in this field.

III. METHODOLOGY

Figure 1 illustrates the data flow of our Machine Learning model, depicting the sequential progression of data throughout the training and prediction processes. Within the realm of Machine Learning, data flow pertains to the systematic passage of data through a model. The data flow commences with data pre-processing, encompassing activities such as data cleansing, normalization, and transformation, aiming to prepare the raw data for compatibility with the machine learning model. Subsequently, the pre-processed data is partitioned into distinct datasets: training, validation, and testing. These datasets are leveraged to train the model and assess its performance.

1. Data Collection: The process of data processing involves the application of diverse techniques to analyze and interpret data effectively. In the context of Bengaluru's real estate properties, we procured data from multiple trustworthy real estate websites. This data encompasses various attributes such as location, square footage, built-up area, and others. Acquiring structured and categorized quantitative data is paramount, as it enables the utilization of machine learning techniques for analysis. Prior to embarking on any machine learning research, it is crucial to source data from reliable channels and validate the dataset's integrity. Without a valid dataset, the analysis of data would lack significance. Hence, meticulous attention must be given to the collection and validation of data to ensure the generation of accurate and dependable outcomes.
2. Data Preprocessing: Data preprocessing plays a pivotal role in data analysis, encompassing crucial tasks such as cleaning and transforming raw data to ensure its suitability for subsequent analysis. During the cleaning process, it is common to encounter missing values or outliers, which can adversely impact the quality of the analysis. To address these challenges, a range of techniques are employed. One prominent approach is data cleaning, which involves identifying and handling missing values and outliers. In the case of missing values, they can be managed by either eliminating the corresponding rows or imputing them with the average value of the variable. Data preprocessing is an essential step, serving to prepare the dataset for subsequent application of machine learning and other data analysis techniques.
3. Model Development: To ensure the accuracy and efficacy of a predictive model, it is a common practice to partition the available data into two distinct sets: a training set and a test set. The training set is utilized to train the model, where input variables are coupled with their corresponding variable, which target the model endeavors to predict. In the case of regression training for the model on models, two widely the training set adopted algorithms are the XGBoost regressor and linear regression. These algorithms strive to discern the underlying relationship between the input variables and the target variable. Following the training phase, the model can be applied to the test set. The test set constitutes a separate data collection that the model has not encountered before, and its purpose is to evaluate the model's capacity to generalize novel data instances. The model's performance on the test set is then utilized to assess its accuracy and effectiveness in generating predictions on new, unseen data.Front-End Development: This may include creating a user interface using ReactJS that allows users to input data.
4. Back-End Development: Tofacilitate communication between the front-end and the machine learning model, an integral aspect involves the development of APIs utilizing Django Rest Framework. These APIs serve as a

crucial interface that enables the front-end to interact with the model and acquire predictions based on user input.

5. Location-Based Search: This will allow users to input their location and view properties by integrating mapping API such as Google Maps into the front-end.
6. Chatbot Integration: This will allow users to ask questions about home construction costs and receive responses from the chatbot.

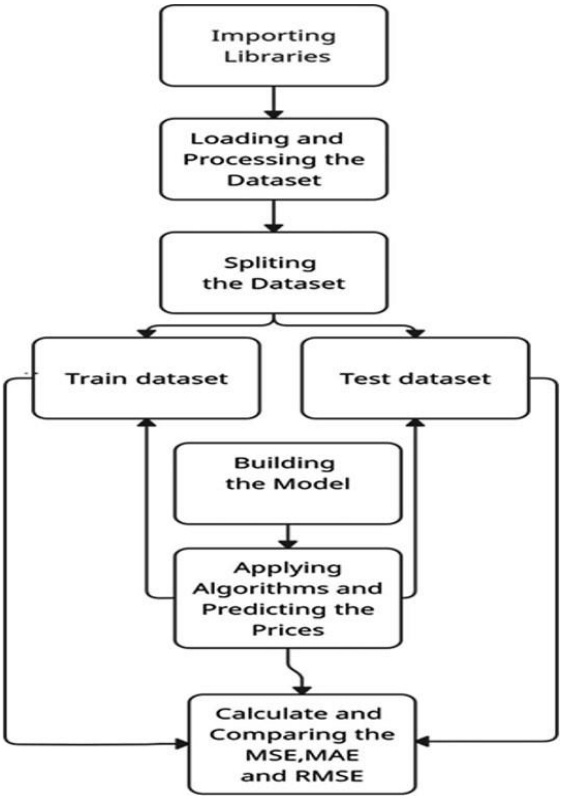

Figure 1: The dataflow diagram of the ML model

Figure 2 displays the workflow diagram, which includes several modules, such as the user module, user view interface, ML module, and database module, as well as an admin module. When a new user visits the website, they can sign up and log in to the system, and user validation will occur. Data is stored in the database, and the admin module can control access to resources for users. Once a user successfully logs in, they are directed to the cost estimation page, where they can input basic information such as location, number of bathrooms, BHK, and square footage. This input goes to the backend of the application using APIs. The front end of the application was developed using Reactjs. In the backend, Django is used to integrate the ML model with the backend, and the ML model receives the inputs and generates a predicted price value that is shown to the user in the frontend. Based on the predicted price, home samples and a GPS map will be displayed.

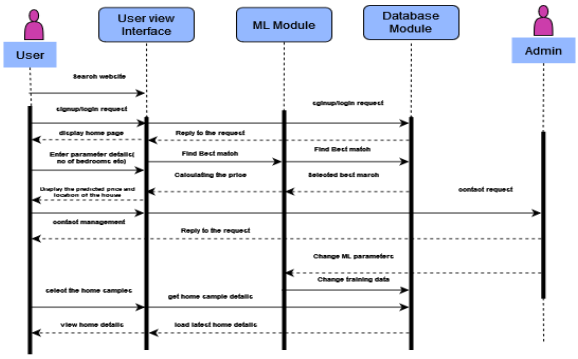

Figure 2: Workflow diagram of the proposed system

Upon completing the construction of the model and achieving successful outcomes, the subsequent step involves integrating it with the user interface (UI). In this regard, React.js is utilized for front-end development, while Django serves as the backend framework. Django, being a web framework, provides a comprehensive range of tools, libraries, APIs, and technologies essential for building web applications. Its features include management effortless route, with a primary focus on seamless integration with Python models. Figure 3 showcases the user interface design of the application. Initially, users are required to log in to the system. Normal users are granted access to the home page, cost estimation page, projects page, and contact page. Conversely, administrators possess additional privileges, enabling them to oversee user management and control the system's workflow.

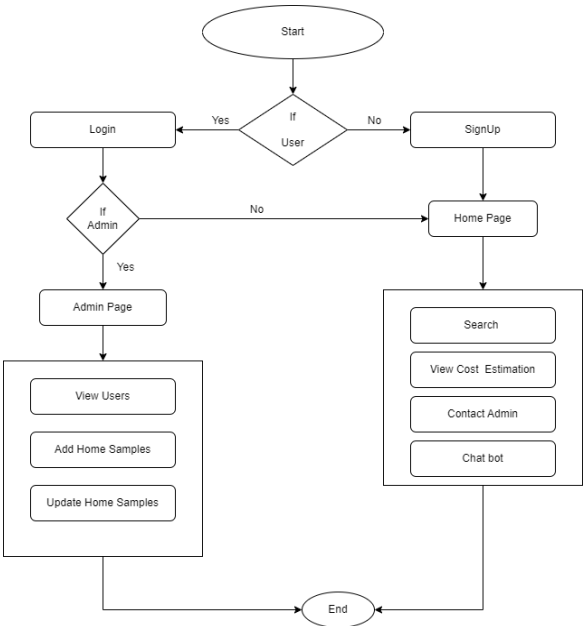

Figure 3: The user interface design of the proposed system is depicted

IV. RESULTS AND ANALYSIS

The primary objective of the project was to utilize machine learning algorithms for predicting home cost estimations. The analysis of the five algorithms, conducted with varying training and testing sizes, is presented in Table 2.

Table 2: Analysis of the five algorithms

Training Size	Testing Size	Accuracy (%)				
		LR	Lasso	Ridge	GBR	XG Boost
90%	10 %	84.74	81.38	84.74	81.51	84.42
80%	20%	85.04	81.93	85.04	82.10	85.65
70%	30%	82.97	80.29	82.97	76.65	78.77
60%	40%	85.10	81.99	85.15	83.36	84.76

Figure 4 shows the accuracy of the five algorithms, with a train set size of 90% and a test set size of 10%, using a bar graph.

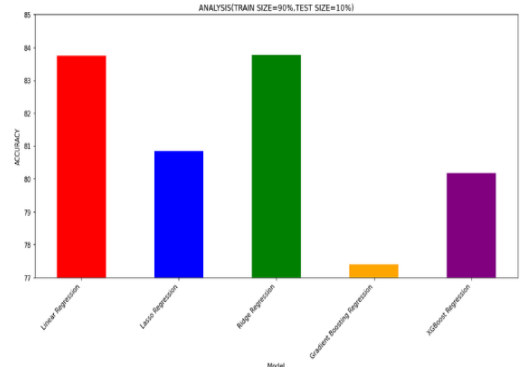

Figure 4: Graph Analysis of the first set

A bar graph, denoted as Figure 5, illustrates the accuracy of the five algorithms with a train set size of 80% and a test set size of 20%.

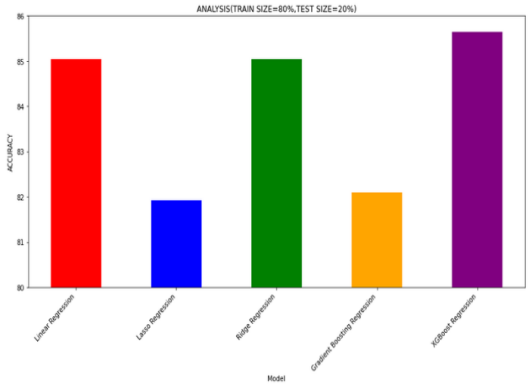

Figure 5: Graph Analysis of the second set

The accuracy of the five algorithms is presented in Figure 6 as a bar graph, with a train set size of 70% and a test set size of 30%.

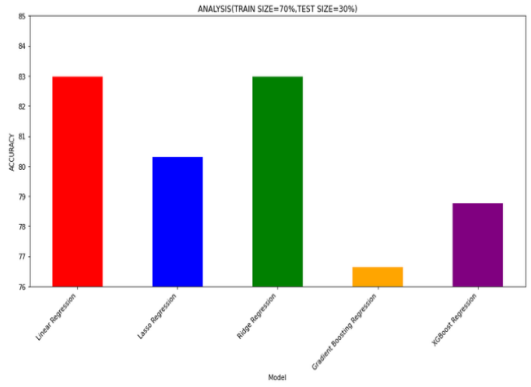

Figure 6: Graph Analysis of the third set

A bar graph, represented as Figure 7, displays the accuracy of the five algorithms with a train set size of 60% and a test set size of 40%.

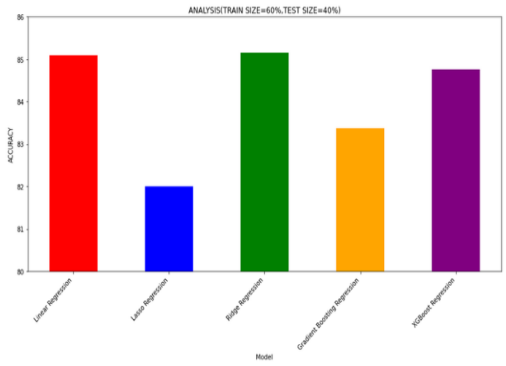

Figure 7: Graph Analysis of the fourth set

Figure 8 is the SignUp page for the users who use this application.Users can SignUp using Google or they can fill the details manually.

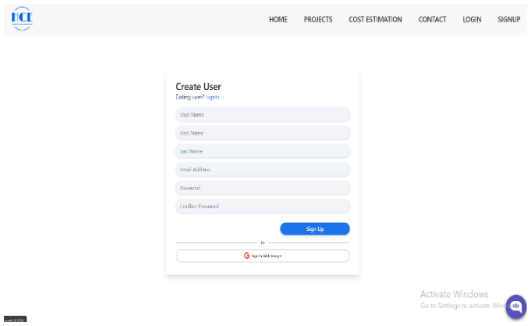

Figure 8: SignUp Page

Figure 9 is the SignIn page for the users who use this application.

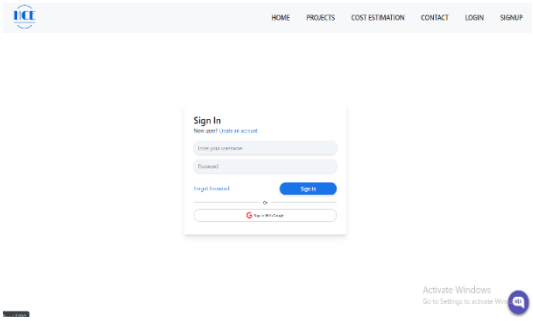

Figure 9: SignIn Page

Figure 10 shows the home page of the website.

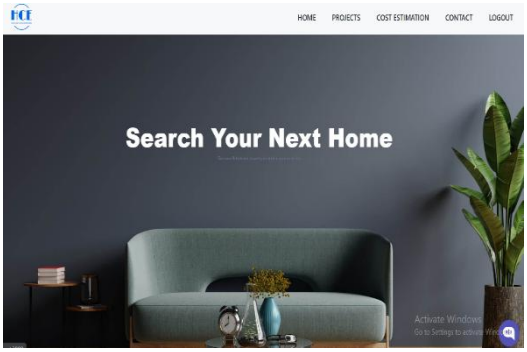

Figure 10: Home Page

Figure 11 is the input page. Here, the user will enter the parameters. Four features are present. These features are the ones that are responsible for the result.

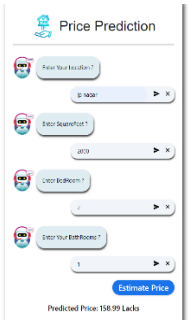

Figure 11: Cost estimation chatbot page

Figure 12 is the result page where users can see home samples based on predicted price range.

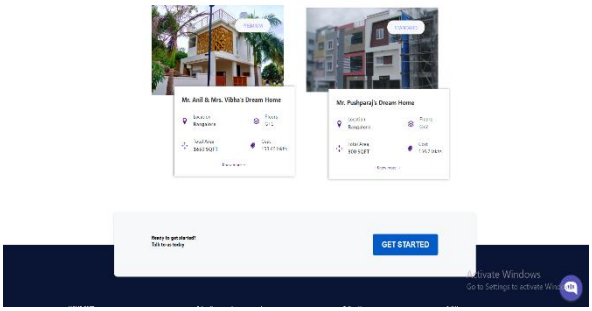

Figure 12: Home samples based on predicted price

Figure 13 shows the chatbot which shows the data based on user input.

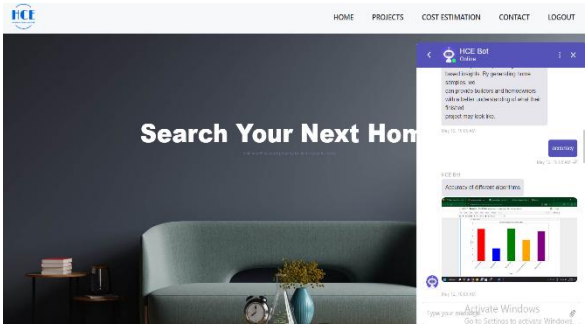

Figure 13: Chatbot

Figure 14 shows the projects page where the recently done projects displaced.

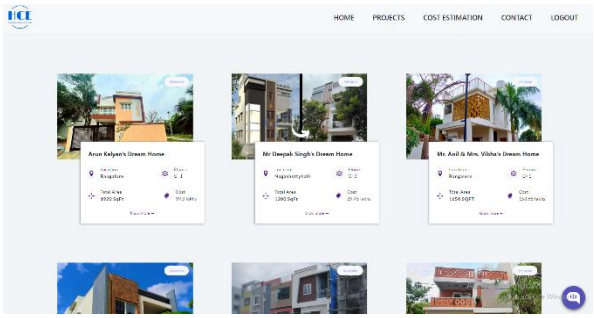

Figure 14: Projects page

Figure 15 shows the description of the home samples shown in the projects.

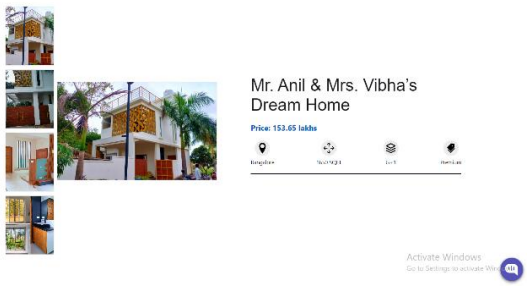

Figure 15: Description about home samples

Figure 16 shows the contact page where user can write their queries.

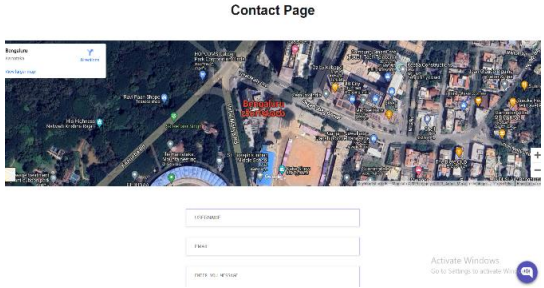

Figure 16: Contact page

CONCLUSION

In recent years, the application of machine learning (ML) techniques has shown significant promise in the field of home cost estimation. ML models consider various factors, such as location, size, age, and property features, to generate more accurate predictions compared to traditional methods. Several algorithms, including linear regression, decision trees, and neural networks, have been explored for ML-based home cost estimation. The selection of an appropriate algorithm depends on the specific requirements of the problem, considering the strengths and weaknesses of each algorithm. The quality and availability of data play a role in the performance of ML critical models. Effective data pre-processing and feature engineering techniques can extract essential information from raw data and improve the accuracy of the models. The utilization of ML in home cost estimation has the potential to revolutionize the real estate industry by providing more reliable and accurate cost estimates. However, further research is needed to address challenges related to data quality and model interpretability.

Future research and development in the field of home cost estimation using machine learning can focus on various areas. These areas include integrating additional data sources, such as economic indicators and demographic data, to enhance the accuracy of the models. Developing more interpretable models will increase trust and confidence in ML-based approaches. Exploring different ML techniques, such as ensemble methods and deep learning, can further improve the performance of home cost estimation models. Incorporating temporal data to capture market trends and changes over time is another avenue for exploration. Evaluating the impact of machines on the learning real estate industry will provide valuable insights. By addressing these areas, researchers and real estate professionals can advance the accuracy of home cost estimation models and gain a better understanding of how machine learning will shape the industry in the future. These advancements have the potential to revolutionize the way home cost estimation is conducted and deliver more reliable and accurate cost estimates, ultimately benefiting the real estate industry.

REFERENCES

[1] Thuraiya Mohd, Suraya Masrom, Noraini Johari, "Machine Learning Housing Price Prediction in Petaling Jaya, Selangor, Malaysia ", International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-2S11, 2019.

[2] SayanPutatunda, "PropTechfor Proactive Pricing of Houses in Classified Advertisements inthe Indian Real Estate Market"

[3] B. Balakumar, P.Raviraj, S.Essakkiammal ,"Predicting Housing Prices using Machine Learning Techniques".

[4] M Thamarai, S P Malarvizhi, " Home cost Prediction Modeling Using Machine Learning",International Journal of Information Engineering and Electronic Business (DJIEEB), VoL12, No.2, pp. 15- 20, 2020. DOI: 10.5815/ijieeb.2020.02.03.

[5] Zulkifley, Nor & Rahman, Ibrahim, Ismail. (2020). Home cost Prediction using a Machine Learning Model: A Survey of Literature. International Journal of Modern Education and Computer Science. 12. 46-54. 10.5815/ijmecs.2020.06.04.

[6] Kuvalekar, Alisha and Manchewar, Shivani and Mahadik, Sidhika and Jawale, Shila, Home cost Forecasting Using Machine Learning (April 8, 2020). Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST)2020.

[7] Byeonghwa Park, Jae Kwon Bae "Using machine learning algorithms for housing priceprediction: The case of Fairfax County, Virginia housing data". 2017

[8] Akshay Babu, Dr.Anjana S Chandran, "Literature Review on Real Estate Value PredictionUsing Machine Learning", International Journal of Computer Science and MobileApplications, Vol: 7 Issue: 3, 2019

[9] M. Venkataraman, V. Panchapagesan, and E. Jalan, 'Does Internet search intensity predict home costs in emerging markets?A case of India,'' Property Manage., vol. 36, no. 1, pp. 103–118, Feb. 2018, doi: 10.1108/PM-01-2017-0003.

[10] C.R.Madhuri, G.Anuradha, and M.V.Pujitha, ''Home cost prediction using regression techniques: A comparative study,'' in Proc. Int. Conf. Smart Struct. Syst. (ICSSS), Mar. 2019, pp. 1–5.

[11] T. D. Phan, ''Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia,'' in Proc. Int. Conf. Mach. Learn. Data Eng. (iCMLDE), Dec. 2018, pp. 35–42.

[12] A. S. Temür, M. Akgün, and G. Temür, ''Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models,'' J. Bus. Econ. Manage., vol. 20, no. 5, pp. 920–938, Jul. 2019.

[13] L. Yu, C. Jiao, H. Xin, Y. Wang, and K. Wang, ''Prediction on housing price based on deep learning,'' Int. J. Comput. Inf. Eng., vol. 12, no. 2, pp. 90–99, 2018.

[14] O. Poursaeed, T. Matera, and S. Belongie, ''Vision-based real estate price estimation,'' Mach. Vis. Appl., vol. 29, no. 4, pp. 667–676, May 2018.

[15] Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick SiowMong Goh - "A hybrid regression technique for home costs prediction" 2017, IEEE.