House Price Predictions & Area based Classification Using Machine Learning

Rahul kumar J, Vaseegarapandian P, Vishnu V, Euodial A

U.G. Student, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli.

U.G. Student, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli.

U.G. Student, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli.

Assistant Professor, Department of Computer Science and Engineering, , Francis Xavier Engineering College, Tirunelveli.

ABSTRACT: This paper presents a machine learning-based approach for house price prediction, aiming to aid buyers in making informed decisions in the real estate market. It is crucial to emphasize that the real estate business is competitive, as it is estimated that 87% of buyers have recently relied heavily on agents or agents for housing purchase guidance. As a result, there is a high demand for accurate pricing predictions, the information that can help buyers choose the best deals. Being aware about the budgeting and marketing strategies as a home buyer, this study uses a Random Forest algorithm as a tool to predict house prices as a result of the data presented to it. By developing a web application dynamic content which is powered by machine learning, buyers can easily want to know expected house prices. As a result, it assists them in exploring this market easily plus making clear investment decisions.

KEYWORDS: Predictions, Price calculator, Linear Regression,, machine learning.

I. INTRODUCTION

The difficulty in predicting the future house prices has an important place in real estate transactions - from the buyers to the sellers everyone seeks trusted resources to guide their final conclusions about what they pay or sell. In fact, in recent years, there has been a significant change in the current land property transaction process with almost 87%, of sales being done through real estate agents and brokers. This spike accentuates the significance of utilizing the most advanced technologies which can include machine learning (ML) or Internet of thing (IoT) that can improve the pricing of the houses and customize the home searching experience for your clients.

Though first-time home buyers are constantly engaged in tackling an array of challenges that come with finding a suitable place in the market, such as balancing the budget and changing the market environment, remaining positive becomes the key in their pursuit of finding the perfect house. While there are huge investment stakes associated with real property, there is an urgent desire for tools and methodologies that can produce estimates of correct and immediate prices so that decision making can be taken from a sound standpoint.

This project aspires to build an AI-based house price prediction system by making use primarily of the knowledge and capabilities of the efficient Random Forest algorithm. Random Forest, one of the most powerful and efficient ensemble learning methods, has proven to be a capable model in capturing complex dataset patterns and filtering out valuable information from it that can be later used for accurate predictions across many areas of application.

With the aid of the machine learning technology, which is of a predictive nature, the project aims to give sellers a user-friendly web application that will be role of enabling the computation of house prices. The technology development brings the integration of sophisticated algorithms and data-driven intelligence. The users can get specific pricing estimates based on the selected details.

Our project is also attempting to create a hybrid model of selling real estate that combines the best features of the traditional real estate practices, and novel technological solution to allem the growing demands of the modern buyers. We hope to achieve this by offering property data driven predictions which are not only

accurate but also display transparency and promote efficiency in the market. By this, we get the trust and loyalty of the property buyers an sellers.

In a world where real estate is shaping up faster through devices, locations and platforms, it feels right to introduce a machine learning-based house price prediction system, which would be a relevant and timely effort. The latest technology will be utilized to give data and info what buyers need to do the better and faster deals. They can buy or rent a suitable house/ apartment because of tools working in the background.

Further to that, the building of a system to predict house prices, with machine learning technology being its backbone, goes beyond the abilities of the average house buyers to create a wider impact on different actors within the real estate sector. An agent or broker would be of great help in this case, as accurate, real time price assessments will ensure the agents offer better advice to their clients and conclude transactions with assurance. Through the medium of machine learning algorithms, the experts of this industry will be able to avail a deep level of understanding about the market trends and dynamics, making it possible to predict how it might be shifting and thus optimizing the use of newly emerging opportunities.

The introduction of machine algorithms is able to offer better prediction for house prices which is also additional to innovation as well as efficiency within the wider real estate sector. By adopting the use of data-driven analysis for the tasks of property evaluation, investments and development, companies and organizations will be able to smooth off the production process, reduce risks and find new opportunities for resource allocation. This fundamental shift to data-driven decision-making, through the use of emerging technology as a tool, has the power to transform age-old practices and create new and more streamlined value-creation processes in the real estate sector.

In which the process of establishing a machine learning-based forecasting system around house prices illustrates, the fact of technological innovation proliferation in housing market is evidence of the broader societal influence. the dissemination of correct pricing information and the empowering of consumers by market participants with greater openness, our project gives the share of voice in the real estate market to the people. One of its main objectives is to achieve these goals by increasing transparency and by making access to decision-making processes more equitable and inclusive in real estate transactions, thus making the housing ecosystem more dynamic and strong for the benefit of individuals and their communities.

II. DATASET

The dataset utilized in our house price prediction project is a comprehensive collection of real estate data spanning various attributes and features pertinent to property valuation and pricing.It includes many aspects that affect home prices including property features, neighbourhood and surroundings, economic variables like unemployment and inflation, and housing market movements.

Important elements of the dataset are such as the property size, number of bedrooms and toilets, gross extraction, the status of the amenities and its state, which is the entirety of the house. Among them, data concerning elements of the physical condition, layout, and specifications of this property are imperative as they would be among the factors that will determine its market value.

Moreover, the dataset contains geographical data that covers the neighbourhood characteristics of demographics, proximity to utilities and services, school district ratings, and the levels of crime. Such location features as the proximity to transit, green spaces, or entertainment facilities give detailed information about the feeling of this place. This implies that specific areas of the city may be more expensive than others depending on location features.

In addition, the economy parameters and market movement are inclusive to the dataset for giving consideration to macroeconomic factor on the real estate field. They can comprise of factors like interest rates, unemployment, average listing and sold property price as well as previous housing inventory. Our model takes into account this volatile nature of the housing market and has the grit to make solid predictions by considering these factors.

Moreover, the data pool is carefully curated and refined which adds to data accuracy and consistency. This is segmented for activities including filling missing records, encoding of categorical variables, and standardization of the numerical features. Through the careful processing we make, we can correct for any biases inherent in the data and the model can be trained to make use of this information appropriately.

Consequently, the dataset can be a uniquely fat and properly varying group of training data for our model that effectively learns house prices. We aim to develop a complex and consistent pricing predictor the essence

of which is to equip home buyers, sellers and respective property professionals with reliable instruments that will allow them to make a well-informed decision in the unstable and competitive housing market.

III. METHODOLOGY

This proposed solution suggests using a multi-faceted method to boost Web application security and user authentication by employing face recognition technologies, personality traits analysis, and voice recognition abilities.

A. Dataset Collection:

Initially, we should have a complete dataset comprising all these characteristics: the surface area, location, amenities, and the economic figures. The dataset processing next involves the missing values handling, categorical variables encoding, numerical features normalization to ensure that the data quality and consistence are well taken care of.

B. Feature Selection:

Following, the dataset is sliced into the components that would have direct influence on how the house prices vary. Additionally feature engineered techniques to form into the way of making new features or transforming existing features may also be useful for the model's performance.

C. Model Selection:

Three different regression algorithms are selected for the house price prediction task: With Linear Regression, Decision Tree Regression, and XGBRegressor. To be reliable, linear regression shall be used because of its simplicity and interpretability, whereas decision tree regression and XGBRegression can be used to represent the data's complicated relationships of nonlinearity.

D. Model Training:

Subsequently, for every regression, the model gets trained on preprocessed data; so for the training set, we employ a portion of the data. In the process of training, the models catch the algorithms and the principal patterns among input features and the final target variable (house value).

D.1. Linear Regression:

• Linear Regression model training calls upon the data in the training set to build the linear equation, which is then used to define the dependence of the dependent variable (house prices) upon the independent variables (features).

• The algorithm training aims at the adjustment of linear model coefficients in order to achieve minimization of difference between the actual house price and predicted values.

• In general, the model emerged by automatically optimizing a cost function, for instance Mean Squared Error (MSE), with OLS or gradient descent in most cases.

• Linear Regression has a simple trainable mechanism and is computationally efficient enough which makes it a convenient option for case where the dataset has many features.

D.2. Decision Tree Regression:

• Training a Decision Tree Regression involves building a binary-tree structure from the training data by node where each node represents a decision based on feature.

• The tree is developed bottom-up through splitting of the data on a feature that gives the best information gain or reduction of impurity at each point of the splitting process.

• At training stage, algorithm determines which the best feature is and the best split point for parsing the data according to the variable of objective (house prices).

• The training process stops according to the rules like: for example, tree depth or size of data samples for each node should be lower than normal.

• A tree-structured Decision Tree Regression could in many situations results in a higher performance of data evaluation because it is capable of capturing complex nonlinear relationships in the data, and is less sensitive to outliers than Linear Regression.

D.3 XGBRegressor (Extreme Gradient Boosting):

• XGBRegressor model is trained via ensemble learning, this philosophy involves training multiple decision trees one after another with the objective of correcting the mistakes of the previous ones.

• During the training phase, the individual decision trees are trained on the residuals (the gap between the actual value and the predicted value) of the previous trees.

• XGBRegressor utilizes gradient boosting which builds an ensemble one model at a time on the existing weak learners (decision trees) in an attempt to reduce a loss function and hence correct the errors of the existing ensemble.

• Training involves automating the tuning process ofvarious hyperparameters like learning rate, tree depth, or the number of estimators to be able to derive the best possible performance of the ensemble model.

• XGBRegressor is famous for its high split conversion rate and resilience because it successfully combines the virtues of decision trees and the gradient boosting way.

E. Model Evaluation and Persistence:

Every model is on the assessment rhythm after training by the use of some metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R2 score which determine the model's performance. These metrics offer various data like how well these models are able to predict the price of house accurately as well as if their predictions are applicable to the entire population.

F. Comparison & Selection:

The comparison is done on the basis of metrics out of the three performances. The model that obtained the most accurate result, which was verified by the evaluation metrics, is chosen that model for house price estimation.

G. XGBRegressor Analysis:

XGBRegressor, showing the highest accuracy compared to the other two models, is therefore further mined to understand and assess its predictive abilitiy. Feature importance assessment is conducted to find the top features that the model XGBRegressor treats as the most significant when it comes of housing estates prices. Views on this study help to understand better which the factors are main that moving the property price.

H. Deployment and Integration:

Last but not least, the author introduces a model and the pertinent preprocessing techniques following deployment and validation and eventually installs an application interface. The application gives users an option to import property attributes in a bid to generate house-prices prediction expertise with XGBRegressor model, which as a model is able to generate house price predictions accurately.

IV. IMPLEMENTATION

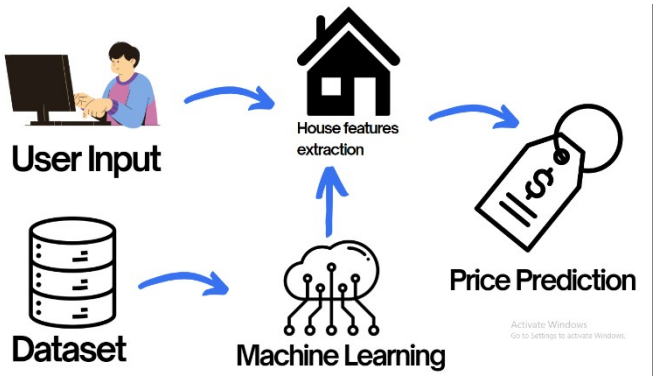

Fig. 1. It represents the flow chart fashion recommendation system.

1. Please ask the consumer for his views on the particularities of the livelihood.
2. A self-learning model will handle the input.
3. Our prediction of price output has been obtained by the model that we have built up with the help of machine learning algorithm.
4. Metrics of performance functions as a form of measuring the accuracy of an algorithm.

V. RESULT

The results of our house price prediction project reveal notable differences in the performance of the three algorithms: Linear Regression,, Decision Tree regression and XGBRegressor. We have observed that the XGBRegressor algorithm is more efficient in modeling the house price and it provides better MAE, MSE and R2 score than other algorithms used. To sum up, XGBRegressor is useful for precise predictions of house prices. The high R2 value which is other than MAE and MSE is the proof of nonlinear regression being superior to linear regression and decision tree regression which do not exhaust all the existing scenarios. Through the utilization of sophisticated machine-intelligence methods, like ensemble learning, XGBRegressor presents the best option for different players in the real estate, such as realtors and potential homeowners, who need accurate and reliable housing market trend predictions.

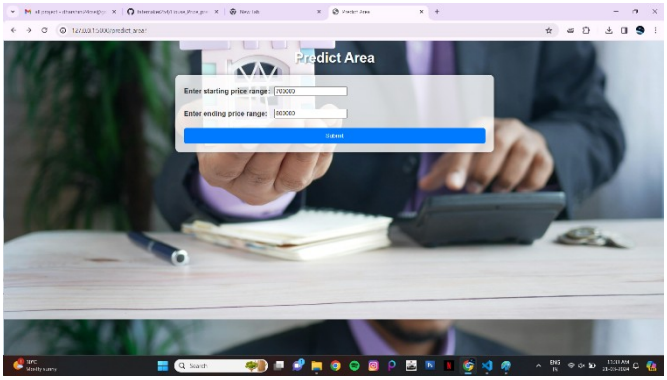

Fig. 2. Output Web App of House Price Predictor

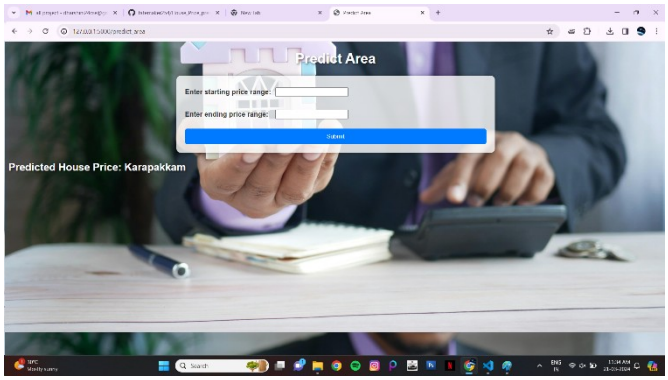

Fig. 3. Output Web App of House Price Predictor

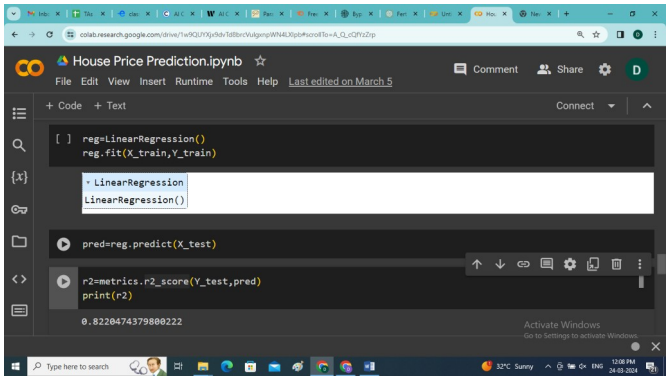

Fig. 4. R-Square value of Linear Regressor – 82%

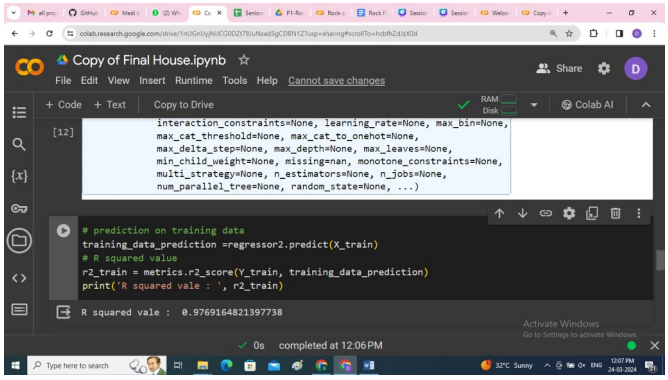

Fig. 5. R-Square value of XGB Regressor is 97%

VI. CONCLUSION

To sum up, our house price forecasting project, which was based on applying machine learning technologies to deal with the challenges of the real estate market in real life, is the benchmark. Through meticulous data preprocessing, feature selection, and model training, we have developed and evaluated three distinct algorithms: The model is built using Linear Regression, Decision Tree Regression, and XGBRegressor Our complete evaluation not only highlighted the reasons for using such algorithms in predicting with a high degree of success, but also revealed other practical applications of then in the field of real estate.

This is in view of the fact that the final results of our project emphasize the point that the type of algorithm employed is crucial. Despite classifying Linear and Regression approaching as easy-to-interpret-inputs, Decision Tree Regression takes a leading position in extracting non-linear correlated data

intimations. Despite the fact, it is the XGBRegressor algorithm that excels, being more accurate and reliable in predicting a housed price compare to other algorithms. XGBRegressor stands out as it has higher R2-value and also performs better among other algorithms because it has a lower error metrics. This implies that the predictions are more reliable and precise.

The effectiveness of XGBRegressor is grounded on the bagging (ensemble) learning technique that blends trees' strength with one of the most powerful machine learning algorithms – gradient boosting. XGBRegressor (eXtreme Gradient Boosting regression) proceeds with the process of iteratively training decision trees in order to correct the errors made by trees in the previous stages. Ultimately, it is able to precisely model complex patterns and intelligently optimize the performance of the model. This highlights the importance of leveraging advanced machine learning techniques to achieve superior predictive capabilities in real-world applications.

Also, as with our project that the transparency and efficiency of the real estate market are boosted by the availability of reliable and accessible price prediction. By equipping buyers, sellers, and real estate experts with truthful indexes, our task creates environments of reliance where correctness and trust reign in the property system.

Looking in advance we can notice potential issues in future research opportunities that concern other factors and elements in order to set an impact on the accuracy and robustness of existing house price prediction models. Furthermore, the implementation of our predictive model into practical environment, e.g. online portals and apps on handheld, could improve effectiveness and usability, therefore more users are likely to take advantage.

In the nutshell project of our house price prediction, we see how machine learning already can and probably will change the face of the real estate the world as we know it. Through the data-based ways and smart algorithms usages, we designed advanced pricing forecasting system that provides customers who operate in housing market with vital benefits. With the advent of technological advancements, we foresee more improvements and innovations are in store for predictive models that will ease the real estate transaction system first in terms of efficiency, transparency and fairness.

REFERENCES

[1] R. Hegde, S. Hegde, S. Kotian and S. Shetty, "Personality classification using Data mining approach", 2019 International Journal of Research and Analytical Reviews(IJRAR), vol. 6, no. 1, March 2019.

[2] Singh Bhawna and Singhal Swasti, "Automated Personality Classification Using Data Mining Techniques", Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020, May 2020.

[3] M. Joshi, S. Fadnaik, A. Shetye and J. Nachankar, "Automated Personality Classification Based On Data Mining Techniques", IEJRD International - Multidisciplinary Journal, vol. 5, no. 5, pp. 5, Jun. 2020.

[4] R. Wald, T. M. Khoshgoftaar, A. Napolitano and C. Sumner, "Using Twitter Content to Predict Psychopathy", 2012 11th International Conference on Machine Learning and Applications, pp. 394-401, 2012.

[5] Keshtkar Fazel, Burkett Candice, Li Haiying and Graesser Arthur, "Using Data Mining Techniques to Detect the Personality of Players in an Educational Game", Studies in Computational Intelligence, vol. 524, pp. 125-150, January 2014.

[6] Y. Saez, C. Navarro, A. Mochón and Pedro Isasi, "A System for Personality and Happiness Detection", 2014 International Journal of Interactive Multimedia and Artificial Intelligence, vol. 2, ISSN 1989–1660.

[7] A. Kumar, A. Gawankar, K. Borge and N. M. Patil, "Student Profile & Personality Prediction using Data Mining Algorithms", International Journal of Advance Research and Innovative Ideas in Education (IJARIIE), vol. 3, no. 2, pp. 2395-4396, 2017.

[8] 5. A. Yata, P. Kante, T. Sravani and B. Malathi, "Personality Recognition using Multi-Label Classification", International Research Journal of Engineering and Technology (IRJET), vol. 05, no. 03, Mar 2018.

[9] 6. Pratama Yudha and Sarno Riyanarto, "Personality classification based on Twitter text using Naive Bayes KNN and SVM", 2015 International Conference on Data and Software Engineering (ICoDSE), pp. 170-174.

[10] 7. S. Başara and O. H. Ejimogu, "A Neural Network Approach for Predicting Personality From Facebook Data", Sage Journal, vol. 11, no. 3, July 2021.

[11] 8. S. Katiyar, H. Walia and S. Kumar, "Personality Classification System using Data Mining", 2020 8th International Conference on Reliability Infocom Technologies and Optimization(ICRITO), pp. 1020-1023, 2020.

[12] Crist´obal Romero and Sebasti´an Ventura, Educational Data Mining: A Review of the State of the Art, vol. 40, no. 6, NOVEMBER 2010.

[13] 11. D. Jurafsky and J. H. Martin, "Naive Bayes Classifier Approach to Word Sense Disambiguation", Computational Lexical Semantics, 2015, [online] Available: http://www.let.rug.nl/nerbonne/teach/rema-stats-methseminar/presentations/Olango-Naive-Bayes-2009.pdf.

[14] I. Cantandir, I. Fernandez-Tobiaz, A. Belllogin, "Relating personality types with user preferences in multiple entertainment domains," EMPIRE 1st Workshop on Emotions and Personality in Personalized Services, 2013.

[15] Aleksandar Kartelj, Vladimir Filipović, Veljko Milutinović, Novel approaches to automated personality classification: Ideas and their potentials.

[16] R. Wald, T. M. Khoshgoftaar,A. Napolitano Using Twitter Content to Predict Psychopathy.

[17] Fazel Keshtkar, Candice Burkett, Haiying Li and Arthur C. Graesser,Using Data Mining Techniques to Detect the Personality of Players in an Educational Game.

[18] Yago Saez , Carlos Navarro , Asuncion Mochon and Pedro Isasi, A system for personality and happiness detection.

[19] J. Golbeck, C. Robles, K.Turner, "Predicting personality with social media," In CHI'11 Extended Abstracts on Human Factors in Computing Systems, pp. 253-262, 2011

[20] D. Zeng, H. Liu, F. Zhao, S. Ge, W. Shen, and Z. Zhang, Proposal Pyramid Networks for Fast Face Detection, vol. 495. Amsterdam, The Netherlands: Elsevier, 2019, pp. 136–149.

[21] Y. Xu, W. Yan, G. Yang, J. Luo, T. Li, and J. He, ''Centerface: Joint face detection and alignment using face as point,'' Sci. Program., vol. 2020, pp. 1–8, Jul. 2020, doi: 10.1155/2020/7845384.

[22] X. Huang, W. Deng, H. Shen, X. Zhang, and J. Ye, ''PropagationNet: Propagate points to curve to learn structure information,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 7263–7272, doi: 10.1109/cvpr42600.2020.00729.

[23] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, ''Deep high-resolution representation learning for visual recognition,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 10, pp. 3349–3364, Oct. 2021, doi: 10.1109/TPAMI.2020.2983686.

[24] H. Du, H. Shi, D. Zeng, X.-P. Zhang, and T. Mei, ''The elements of endto-end deep face recognition: A survey of recent advances,'' ACM Comput. Surv., vol. 54, pp. 1–42, Jan. 2022, doi: 10.1145/3507902.

[25] Y. Liu, X. Tang, J. Han, J. Liu, D. Rui, and X. Wu, ''Hambox: Delving into mining high-quality anchors on face detection,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 13043–13051, doi: 10.1109/CVPR42600.2020.01306.

[26] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, ''Joint face detection and alignment using multitask cascaded convolutional networks,'' IEEE Signal Process. Lett., vol. 23,