Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10575019

Enhanced Real Estate Forecasting using Decision Tree with Hyperparameters Tuning

Soumojit Roy
Dept. of CSE (AIML)
Haldia Institute ofTechnology
Haldia, W.B, India
Soumojitroy403sr@gmail.com

Subir Gupta
Dept. of C.S.E.(AI & ML)
Haldia Institue of Technology
Haldia, W.B. India
subir2276@gmail.com
ORCID 0000000209410749

Pronita Mukherjee
Dept. of CSE.
Gargi Memorial Institute of Technology
Baruipur, W.B. India
pronita.csegmit@jisgroup.org

***Abstract*—This research study analyzes the efficacy of using decision trees enhanced by hyperparameter tuning for real estate forecasting, a domain marked by complexity and dynamic nature. This study explores the integration of decision tree algorithms within real estate analytics, highlighting its adaptability and depth in addressing the sector's challenges. The decision tree's ability to determine non-linear relationships and diverse data, helps to accurately predict real estate market trends and property values. Hyperparameter tuning emerges as a critical process that refines the decision tree model, optimising its parameters to align with the unique attributes of real estate data, thereby enhancing predictive accuracy and model generalizability. This study demonstrates various methodologies, from improving market trend analysis and property valuation to customizing models for commercial real estate insights. This study represents these insights, illustrating how decision trees with hyperparameter tuning offer a robust analytical tool for real estate forecasting, providing stakeholders with a understanding of market dynamics and informed decision-making.**

Keywords— Decision Tree, Hyperparameter tuning, Real estate, Forecasting model

I. INTRODUCTION

The real estate sector's substantial economic implications and dynamic nature necessitate precise and adaptable predictive models to navigate its complexities effectively[1]. Traditional models often struggle to accommodate the sector's multi-faceted nature, leading to less reliable forecasts[2]. This research study introduces an advanced predictive modelling approach, integrating decision trees with hyperparameter tuning, to address these challenges in real estate forecasting. The aim is to provide a comprehensive tool that enhances forecast accuracy, adaptability, and interpretability, supporting informed decision-making across the real estate industry[3]. The inherent complexity of real estate markets arises from their susceptibility to various influencing factors, including but not limited to economic cycles, interest rates, demographic shifts, and policy changes[4]. These elements introduce volatility and non-linearity that standard predictive models, which often rely on linear assumptions or fail to capture complex interactions between variables, cannot adequately address[5].

The introduction of decision trees in this context marks a pivotal advancement. Decision trees are versatile and can model complex patterns in real estate data without needing assumptions about how data is distributed, making them practical for detailed forecasting. However, fine-tuning decision trees with hyperparameter tuning can significantly boost their accuracy. This method fine-tunes the model better to match the unique aspects of real estate data, significantly improving its effectiveness. Hyperparameter tuning adjusts critical settings, such as tree depth and splitting points, to boost the model's accuracy and prevent overfitting.

Overfitting happens when a model picks up random noise instead of valuable patterns, which can harm an unpredictable market like real estate, reducing the model's effectiveness on new data [6]. Fine-tuning the hyperparameters enhances the decision tree's ability to generalise, thus increasing the accuracy and relevance of its predictions across various market conditions. This improved model benefits everyone in the real estate sector, including investors, developers, policymakers, and financial institutions. For investors and developers, it provides more accurate forecasts that help them make informed decisions about property investments, thereby boosting profitability and reducing risks. Policymakers can use these insights to create more effective housing policies that align with market needs. Financial institutions gain a reliable tool for property valuation and risk assessment, aiding in more secure lending and investment choices. Unlike more complex models, decision trees remain easy to understand and interpret [7][8]. Each decision point in the tree is direct and based on observable data features, allowing stakeholders to understand and trust the model's predictions. This transparency is invaluable in the real estate sector, where decisions often involve substantial financial stakes and require justification from various parties, including investors, regulators, and the public [9].

The integration of decision trees with hyperparameter tuning presents a significant methodological advancement in the field of real estate forecasting. It addresses the critical need for a predictive model that is accurate and adaptable to the sector's complicated dynamics while remaining transparent and interpretable for various stakeholders [10][11]. The demand for sophisticated, reliable, and understandable predictive tools will undoubtedly grow as the real estate market evolves, influenced by many economic, social, and environmental factors. This research offers a methodological framework that meets this demand and sets a new standard for predictive analytics in the real estate sector, promising enhanced decision-making capabilities and strategic insights for all market participants. The rest of the paper has been structured, followed by section II, described as a literature review, and section III, which discusses the proposed methodology. Sections IV and V elaborate on the results and conclusion, respectively.

II. LITERATURE REVIEW

Some research studies have utilized decision trees for real estate forecasting, particularly when augmented by hyperparameter tuning, spans a broad spectrum of approaches and findings, demonstrating the model's adaptability and depth in analysing complex real estate data. This section helps to analyse the progressive integration of decision tree algorithms in real estate, each contributing unique insights and methodologies to the field [12]. The initial research sets the stage by introducing decision trees in real estate analysis, emphasising their capacity to navigate the sector's non-linear and diverse data landscape for predicting prices and market trends. Another study analyses the hyperparameter tuning, showcasing its pivotal role in refining machine learning models for enhanced accuracy and applicability in the fluctuating real estate market [13]. A comparative analysis then positions decision trees alongside various Machine Learning (ML) techniques in property valuation, highlighting the model's interpretability and how its predictive quality can be enhanced with hyperparameter adjustments[14].

Another study enhances decision trees by adding geographic data, significantly improving forecast precision and emphasising the model's sensitivity to spatial factors. Exploring hyperparameter optimisation more broadly, an existing research reviews various tuning strategies like grid and randomised search, emphasising their relevance in fine-tuning decision trees for real estate applications[15][16]. A research study broadens the view on AI in real estate, particularly highlighting how tailored hyperparameter adjustments optimise decision trees for real estate pricing challenges. Further, research on commercial properties shows how adjusted decision trees reveal investment opportunities and market trends, providing deep insights into commercial real estate dynamics. These investigations collectively demonstrate the decision tree model's growing significance in real estate analytics, driven by hyperparameter tuning to boost predictive accuracy and flexibility. They illustrate a trend towards methodological innovation and varied applications, showing how these tools are continuously adapted for the detailed demands of real estate forecasting. This body of research forms a strong foundation for future work, pointing towards expanded uses of decision trees in real estate to deliver richer insights and enhanced predictive power in this evolving field [17].

III. METHODOLOGY

The proposed methodology is developed to enable predictive modelling, ensuring the robustness and credibility of research findings. This comprehensive framework seamlessly integrates traditional statistical analysis with cutting-edge machine learning techniques, providing an in-depth and insightful review into predictive modelling. The initial phase, data collection through surveys, is critical as it captures a dataset representative of a diverse demographic range, thereby enriching subsequent analytical efforts with high-quality and relevant data. The data cleaning process eliminates anomalies, establishing a foundation for future analysis and informing the selection of variables for complex predictive modelling.

Applying the Yeo-Johnson transformation during the power transformation phase ensure the analytical precision. This step normalizes data distributions and enhances the interpretive validity of conclusions, addressing issues such as skewness and preparing the dataset for advanced predictive modelling. In the predictive modeling with decision tree regressor phase, the study leverages the robust capabilities of this algorithm to uncover complex, non-linear relationships within the data. The inherent flexibility of this model with linear constraints, makes it particularly suitable for capturing complex predictive dynamics. Hyperparameter tuning with RandomizedSearchCV marks a refinement phase, employing a strategic and efficient approach to optimising the model's settings through a randomised parameter search. This method effectively explores the hyperparameter space, significantly enhancing the model's predictive precision and robustness. The subsequent model validation and serialisation phase involves critically evaluating the refined model against an unseen test dataset, using metrics like Mean Squared Error (MSE) and R-squared to provide a assessment of the model's predictive power and real-world applicability. Post-validation, the serialisation of the model ensures its availability for future academic and practical applications, highlighting the methodological contribution to the field. Figure 1 illustrates the methodology diagram, and Table 1 presents the pseudocode of the proposed methodology, providing a clear and detailed guide to the research process.

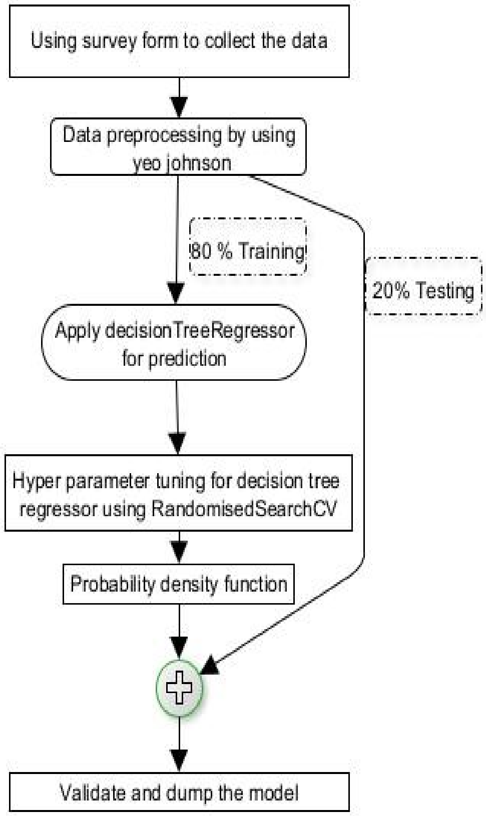

Figure 1: Methodology Diagram

This methodological design not only supports the study's findings in terms of reliability and validity but also stands as a guiding framework, providing a replicable model for future research in predictive analytics.

This study analyses the efficacy of Decision Tree regressions, mainly focusing on how hyperparameter tuning can significantly enhance their predictive capabilities. The study investigation centres around two models: one with a fixed maximum depth of 5 and another optimised through RandomizedSearchCV. The evaluation of these models pivots on three fundamental metrics: Mean Absolute Error (MAE), and Mean Squared Error (MSE), which collectively offer a multi-faceted view of model performance. Initially, the model with a predetermined depth exhibited an MAE of 0.3046, suggesting that, on average, the model's predictions deviated from actual values by this margin. An MSE of 0.1686 pointed to more significant errors being disproportionately penalised, hinting at potential improvement areas.

Table 1:- Pseudocode for the proposed methodology

1: data ← collect data(surveyForm)
2: analyzeUnivariate(data)
3: analyzeBivariate(data)
4: data_transformed ← yeoJohnson(data) // Apply Yeo-Johnson: data_transformed_i = f(data_i)
5: model ← DecisionTreeRegressor() // Initialize model: f(X, Θ) → Y
6: params_dist ← {param1: dist1, param2: dist2, ...} // Define hyperparameter distributions
7: optimal_params ← RandomizedSearchCV(model, params_dist, n_iterations) // Optimize: Θ_opt = argmin(Λ(Θ))
8: model.setParameters(optimal_params)
9: model_trained ← train(model, data_transformed) // Train: minimize $\sum\left(y - f(x, \Theta\_\text{opt})\right)^2$
10: validation_score ← validate(model_trained, data_validation) // Validate: score = $\sum\left(y - f(x\_val, \Theta\_\text{opt})\right)^2$
11: dump(model_trained) // Serialize model for deployment

IV. RESULTS & DISCUSSION

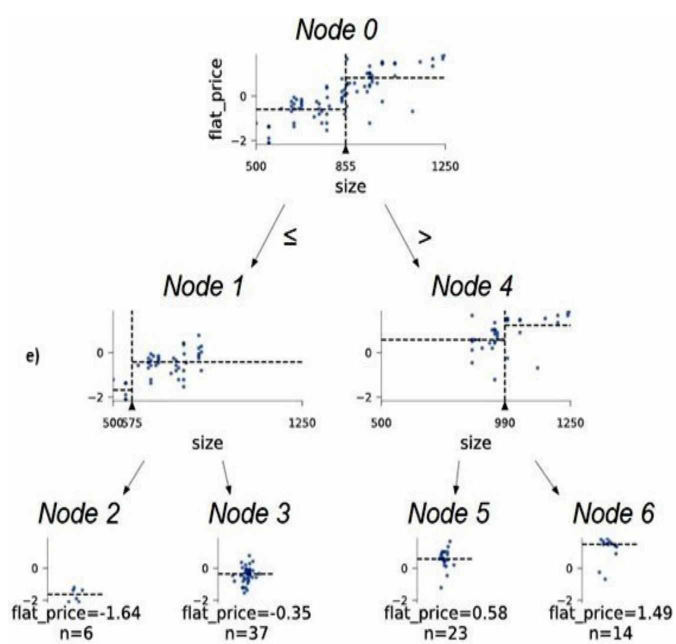

Figure 2:- Decision tree after hyperparameter tuning (using randomizedSearchCV)

Figure 2 shows the decision tree after hyperparameter tuning. The R^2 score of 0.8393 initially indicated that the model's input features could explain approximately 83.93% of the output variance. After applying hyperparameter tuning via RandomizedSearchCV, there was a marked improvement across all performance metrics. The Mean Absolute Error (MAE) decreased to 0.2534, denoting enhanced prediction accuracy. Similarly, the Mean Squared Error (MSE) reduced to 0.1276, reflecting a more consistent performance across various data points and minimizing the impact of larger errors. Most notably, the R^2 score ascended to 0.8711, indicating that the tuned model accounted for approximately 87.11% of the variance in the dataset. This substantial improvement signifies a tighter fit to the data. Table 2 provides a comparative analysis of the results between the decision trees before and after hyperparameter tuning. The study also analyzed the complexities of variable interrelations, utilizing univariate and correlation analyses to highlight key trends and anomalies that influenced model performance..

TABLE 2:- COMPARISON RESULTS OF DECISION TREE WITH HYPER TUNING AND WITHOUT HYPER TUNING

	MAE	MSE	R^2 SCORE
Decision Tree with max depth=5	0.3046	0.1686	0.8393
Decision Tree after hyper-parameter tuning(using randomizedSearchCV)	0.2534	0.1276	0.8711

This study highlights the critical role of precise model tuning in enhancing the predictive accuracy and reliability, aiming to interpret complex technical details for broader accessibility. It enriches academic discussions and offers practical insights on employing decision tree regressors in real estate, promoting a detailed and sophisticated approach for data analysis.

V. CONCLUSION

Integrating decision trees with hyperparameter tuning represents a significant advancement in real estate analytics, enhancing market trend analysis and property value forecasts. This research highlights the efficacy of decision trees in handling complex, non-linear real estate data, while hyperparameter tuning optimizes model performance and reliability. This study highlights the integration of decision trees in real estate forecasting, demonstrating their adaptability and precision. These findings collectively confirm the decision tree's capability to decode the intricate dynamics of the real estate market, with hyperparameter tuning playing a crucial role in improving accuracy and usability. The implications are insightful for various stakeholders: investors and developers gain a powerful tool for strategic decision-making, policymakers can craft precise urban development policies, and financial institutions benefit from accurate property valuations and risk assessments. Future research could explore the synergy between decision trees and other advanced machine-learning techniques, and investigate different hyperparameter tuning methods across various market scenarios. Additionally, applying these models to emerging markets or analyzing the impact of new economic policies could enhance their relevance and effectiveness. This study highlights the current strengths and potential evolution of decision tree models, essential for meeting the growing demand for sophisticated and data-driven analytics in the real estate sector.

REFERENCES

[1] T. Su, H. Li, and Y. An, "A BIM and machine learning integration framework for automated property valuation," J. Build. Eng., vol. 44, no. May, 2021, doi: 10.1016/j.jobe.2021.102636.
[2] H. J. Sheu and Y. C. Wei, "Effective options trading strategies based on volatility forecasting recruiting investor sentiment," Expert Syst. Appl., vol. 38, no. 1, pp. 585–596, 2011, doi: 10.1016/j.eswa.2010.07.007.
[3] S. Paul and P. Sharma, "Forecasting gains by using extreme value theory with realised GARCH filter," IIMB Manag. Rev., vol. 33, no. 1, pp. 64–70, 2021, doi: 10.1016/j.iimb.2021.03.011.
[4] F. Petropoulos et al., "Forecasting: theory and practice," Int. J. Forecast., no. xxxx, 2022, doi: 10.1016/j.ijforecast.2021.11.001.
[5] K. Lisspers et al., "Developing a short-term prediction model for asthma exacerbations from Swedish primary care patients' data using machine learning - Based on the ARCTIC study," Respir. Med., vol. 185, no. February, 2021, doi: 10.1016/j.rmed.2021.106483.
[6] Z. U. Rehman, M. S. Zia, G. R. Bojja, M. Yaqub, F. Jinchao, and K. Arshid, "Texture-based localisation of a brain tumour from MR-images using a machine learning approach," Med. Hypotheses, vol. 141, no. January, p. 109705, 2020, doi: 10.1016/j.mehy.2020.109705.
[7] A. Shokry, S. Medina-González, P. Baraldi, E. Zio, E. Moulines, and A. Espuña, "A machine learning-based methodology for multi-parametric solution of chemical processes operation optimisation under uncertainty," Chem. Eng. J., vol. 425, 2021, doi: 10.1016/j.cej.2021.131632.
[8] R. B. Ghannam and S. M. Techtmann, "Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring," Comput. Struct. Biotechnol. J., vol. 19, pp. 1092–1107, 2021, doi: 10.1016/j.csbj.2021.01.028.
[9] A. O. Anim-Ayeko, C. Schillaci, and A. Lipani, "Automatic blight disease detection in potato (Solanum tuberosum L.) and tomato (Solanum lycopersicum, L. 1753) plants using deep learning," Smart Agric. Technol., vol. 4, no. November 2022, p. 100178, 2023, doi: 10.1016/j.atech.2023.100178.
[10] T. K. Ho, "Random decision forests," Proc. Int. Conf. Doc. Anal. Recognition, ICDAR, vol. 1, pp. 278–282, 1995, doi: 10.1109/ICDAR.1995.598994.
[11] J. R. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/bf00116251.
[12] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer," Ann. Med. Surg., vol. 62, no. January, pp. 53–64, 2021, doi: 10.1016/j.amsu.2020.12.043.
[13] E. Menahem, A. Shabtai, L. Rokach, and Y. Elovici, "Improving malware detection by applying multi-inducer ensemble," Comput. Stat. Data Anal., vol. 53, no. 4, pp. 1483–1494, 2009, doi: 10.1016/j.csda.2008.10.015.

[14] F. Cabitza et al., "Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests," Clin. Chem. Lab. Med., vol. 59, no. 2, pp. 421–431, 2021, doi: 10.1515/cclm-2020-1294.

[15] T. Jefferson et al., "Physical interventions to interrupt or reduce the spread of respiratory viruses: Systematic review," Bmj, vol. 336, no. 7635, pp. 77–80, 2008, doi: 10.1136/bmj.39393.510347.BE.

[16] E. R. Davies, "Face detection and recognition," Comput. Vis., no. March, pp. 631–662, 2018, doi: 10.1016/b978-0-12-809284-2.00021-6.

[17] B. Mondal, D. Chakraborty, N. K. Bhattacherjee, P. Mukherjee, S. Neogi, and S. Gupta, "Review for Meta-Heuristic Optimization Propels Machine Learning Computations Execution on Spam Comment Area Under Digital Security Aegis Region," in Integrating Meta-Heuristics and Machine Learning for Real-World Optimization Problems, P. JanuszKacprzyk, Polish Academy of Sciences, Warsaw, Ed. Springer Nature, 2022, pp. 343–361. doi: 10.1007/978-3-030-99079-4_13.