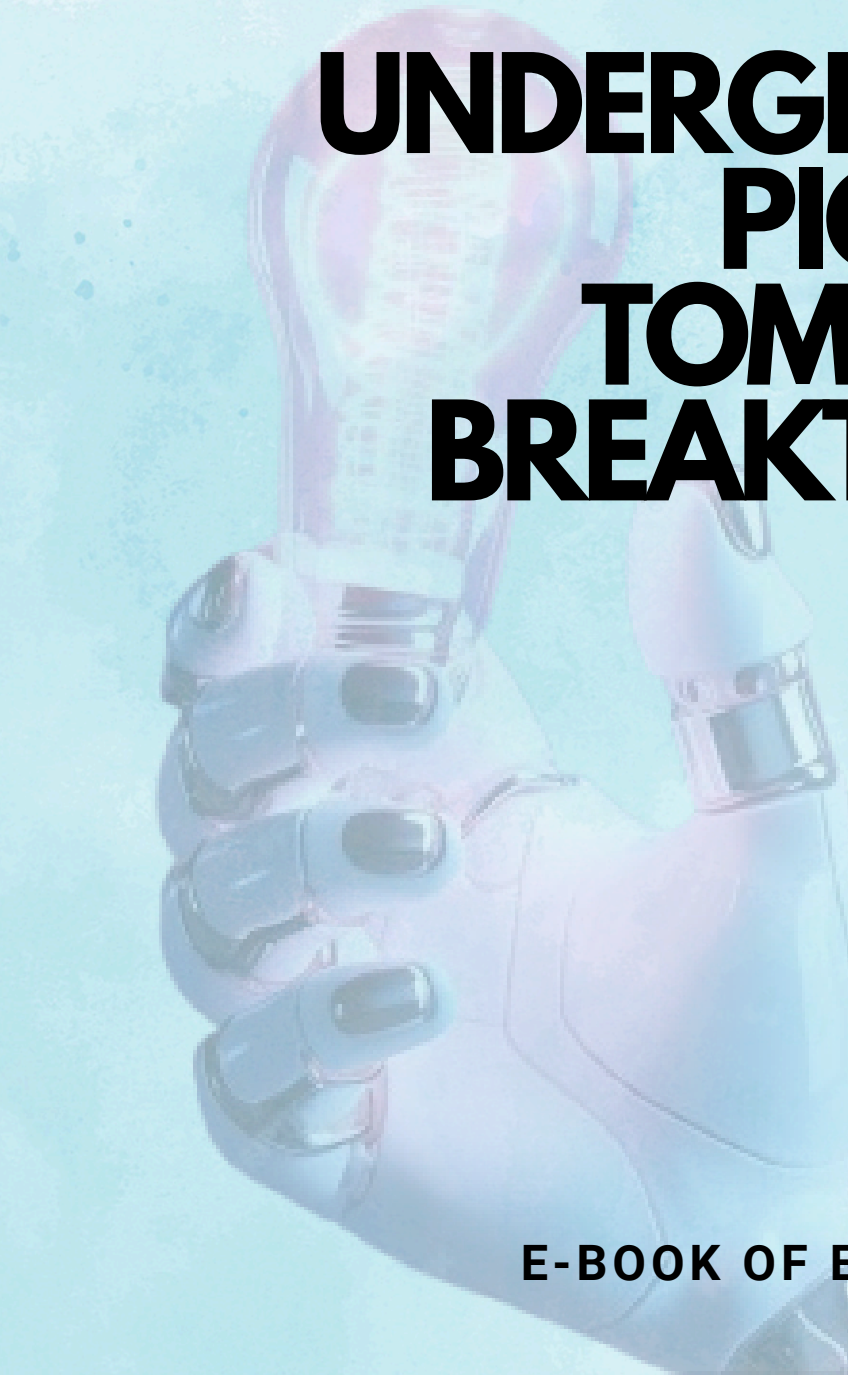
UNIVERSITI TEKNOLOGI MARA

I-URID 2025

INTERNATIONAL UNDERGRADUATE RESEARCH INNOVATION, INVENTION AND DESIGN

UNDERGRADUATES PIONEERING TOMORROW'S BREAKTHROUGH



E-BOOK OF EXTENDED ABSTRACTS

PREDICTIVE MODELING OF PROPERTY PRICE IN MALAYSIA USING TREE-BASED MODELS IN MACHINE LEARNING

Tanushalani A/P Muniandy*

Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob 26300 Kuantan Pahang, Malaysia

*tanushalanimuniandy@gmail.com

ABSTRACT

This study aims to conduct predictive modelling on property price prediction in Malaysia's property market. The market, influenced by economic conditions, supply and demand, and government regulations, is dynamic and multifaceted; hence advanced forecasting methods are essential. Traditional regression models often fail to handle nonlinear and complex relationships. This study applies tree-based machine learning models, namely Decision Tree and Random Forest, for higher accuracy in property price prediction and to provide insight into the main determinants affecting pricing. Data was obtained from the National Property Information Centre, covering transactions across various property sectors from 2001 to 2022. Data cleaning, feature engineering, and data visualization in Tableau were performed to ensure integrity and engagement. Predictive modelling focuses on a comparison between Decision Tree and Random Forest. The objectives are to create an interactive dashboard, develop predictive models, and determine the best tree-based approach for predicting Malaysian property prices. Results indicate that Random Forest outperformed Decision Tree in accuracy, precision, recall, and F1-score. The visualisation dashboard and predictive models will assist stakeholders in policymaking, investment, and housing decisions. This contributes to a more transparent, data-driven, and equitable Malaysian property market, supporting future urban planning and economic strategies.

Keywords: Property price prediction, machine learning, Malaysian property market, data visualization, tree-based models

INTRODUCTION

The Malaysian property market is an important component of the national economy with interaction of economic, social and policy-driven forces. Accurate price prediction is important for homebuyers, sellers, investors and policymakers as it supports decision making within the property market which is often marked by fluctuations and complexity (Shahirah et al., 2021). In recent years, new opportunities have been created with rapid technological advances to address the challenges in facing property valuation and forecasting. Machine learning is able to predict property prices accurately based on amenities, size, location, accessibility and market conditions by leveraging large volumes of data and discovering complex patterns. This study utilises tree-based models and machine learning principles to develop the most accurate and reliable methods of predicting property prices in Malaysia.

The Malaysian property market is becoming more complicated and dynamic without much data-driven insights to guide stakeholders in their decision-making process. Some of the past research mainly highlights the application of machine learning models for property price prediction but lack of focus on real-time analysis to address the evolving market dynamics effectively. Insufficient analysis on property price trends is restraining buyers, investors and policymakers from predicting the market behaviour and hence making proper decisions on property investments. Regression models are generally utilised because of their simplicity for analyzing relationships between variables, but the inability to capture nonlinear and interactive effects seriously impairs their usefulness in property market studies. This emphasises the need to adopt tree-based models to improve property price prediction.

The main objective of this study is to address the challenges of property price prediction in Malaysia through data visualisation and machine learning. The study aims to create an interactive visualisation dashboard that allows users to explore and analyse property price trends and regional differences in Malaysia, to develop predictive models for Malaysian property prices using the Decision Tree and Random Forest models in machine learning, and to determine which tree-based machine learning model provides the highest accuracy in predicting property prices in Malaysia while optimising it for better results.

METHODOLOGY

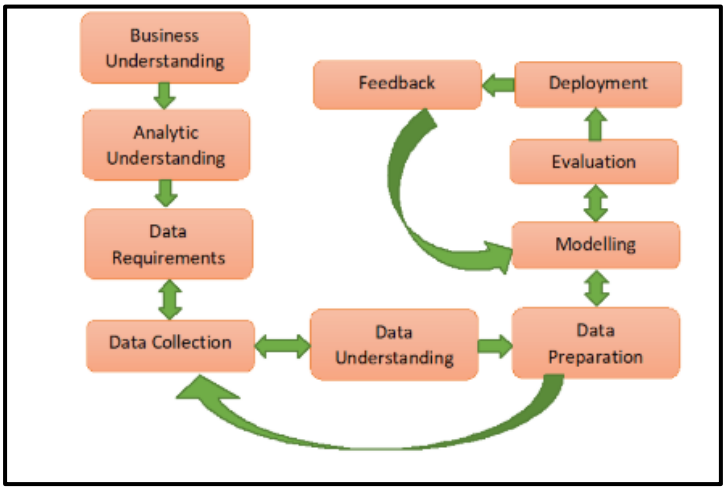


Figure 1: Research Flowchart

The research methodology for this study involves a systematic approach to predict property prices in Malaysia, as outlined in the research flowchart in Figure 3. The process begins with the Business Understanding phase, which identifies the core problem of predicting property prices, analysing influencing factors, and segmenting the property market. This is followed by the Analytic Understanding phase, where business goals are converted into analytical tasks, and the Data Requirements phase, which defines the necessary data sourced from the National Property Information Centre (NAPIC). The next stages include Data Collection, Data Understanding to identify data types and issues, and Data Preparation, which involves cleaning and transforming the dataset. The data cleaning process included restructuring the dataset headers, removing columns with null values, renaming fields for clarity, and handling missing or duplicate entries to ensure data integrity before further analysis.

Once the dataset was cleaned and prepared, predictive modelling was performed using Decision Tree and Random Forest algorithms to capture non-linear relationships and identify significant factors influencing property prices. The models were evaluated using accuracy, precision, recall, and F1-score to ensure robust performance, while hyperparameter tuning was applied to enhance predictive accuracy. Following this, an interactive visualisation dashboard was deployed through Tableau Public to analyse property price trends, regional disparities, and market opportunities. Finally, a Feedback phase collected inputs from stakeholders such as investors, policymakers, and developers to refine the models and dashboards for improved usability and decision-making support.

RESULTS AND DISCUSSION

An interactive dashboard was developed to analyze property price variations across sectors, fulfilling Objective 1 of this study. It allows users to explore sectoral differences, identify price patterns, and drill down into property values by section and year. The dashboard presents data through user-friendly charts and graphs, making fluctuations easier to interpret and supporting stakeholders in property market decision-making. The dashboard can be accessed via the following Tableau Public link:
https://public.tableau.com/app/profile/tanushalani.muniandy5320/viz/MalaysianPropertyMarketOverview/Introduction

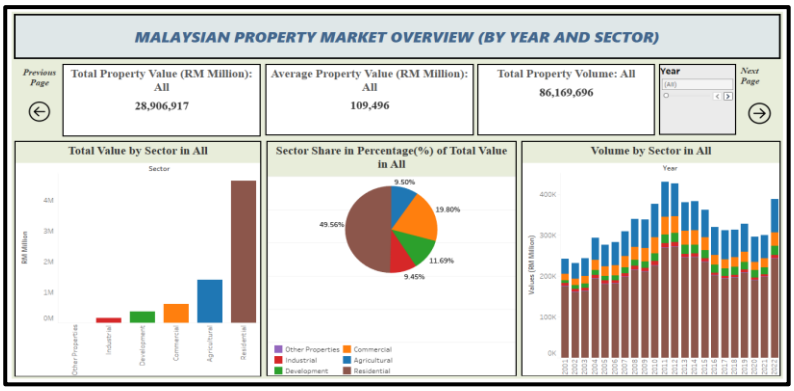


Figure 2: Malaysian Property Market Overview Dashboard

Figure 2 presents a comprehensive dashboard of the Malaysian property market from 2001 to 2022, highlighting key metrics such as total property value, average value, and transaction volume. Over this period, the total property value amounted to RM 28,906,917 million, reflecting the overall scale of transactions, while the average property value of RM 109,496 million serves as a benchmark for market growth. The total of 86,169,696 transactions underscores strong market activity and liquidity. Sector analysis shows that residential properties consistently dominate in both value and share at 49.56%, followed by the commercial sector at 19.8%, with industrial and agricultural sectors making smaller contributions.

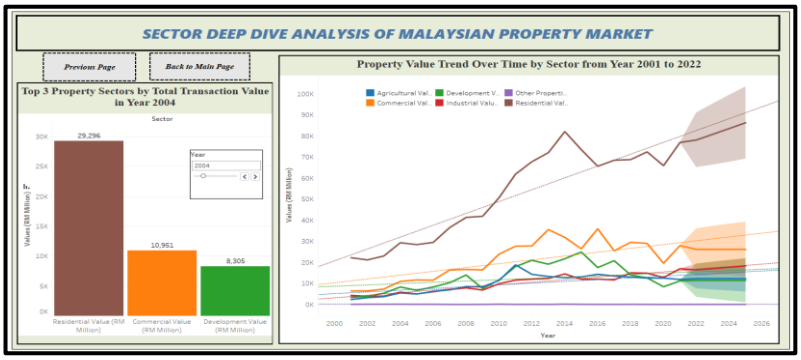


Figure 3: Sector Deep Dive Analysis of Malaysian Property Market Dashboard

The second dashboard in Figure 3 provides a comprehensive analysis of the Malaysian property market, focusing on premier property sectors and their trends over time. It highlights the Top 3 Property Sectors by Total Transaction Value for 2004 and the Property Value Trend Over Time from 2001 to 2022. In 2004, the residential sector led with RM 29,296 Million, followed by the commercial sector at RM 10,951 Million and the development sector at RM 8,305 Million. The trend line from 2001 to 2022 shows residential properties with the highest and most consistent growth, while other sectors such as commercial, industrial, agricultural, and development also display upward trends. Overall, the analysis reflects a general rise in property values across Malaysia over the years.

Table 1: Evaluation metrics of model before and after hyperparameter tuning

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Before Tuning				
Decision Tree	0.40	0.60	0.40	0.48
Random Forest	0.80	1.00	0.80	0.88
After Tuning				
Decision Tree	0.60	1.00	0.60	0.75
Random Forest	0.80	1.00	0.80	0.88

The second objective of this study is to develop predictive models for property prices in Malaysia using the Decision Tree and Random Forest algorithms in machine learning. Both models are developed in a number of steps which involve data preparation, model training, evaluation, and optimization. The Decision Tree model intended to predict Malaysian property price categories based on volume and year of each sector. The model building process involves training the classifier, evaluating predictive performance, and hyperparameter tuning to optimise its accuracy. After tuning the hyperparameters, Decision Tree achieved accuracy at 0.60 with improved precision, recall, and F1 score compared to the initial model, as shown in Table 1.

Meanwhile, the Random Forest model performed well both before and after hyperparameter tuning in predicting Malaysian property price categories based on year and transaction volumes across sectors. It achieved an accuracy of 0.80 with precision of 1.00, recall of 0.80, and F1-score of 0.88, where 80% of predictions were correct. After tuning, the optimised model retained the same performance, confirming its robustness. Compared to the Decision Tree, the Random Forest consistently demonstrated superior accuracy, precision, recall, and F1-score, making it the best-performing model.

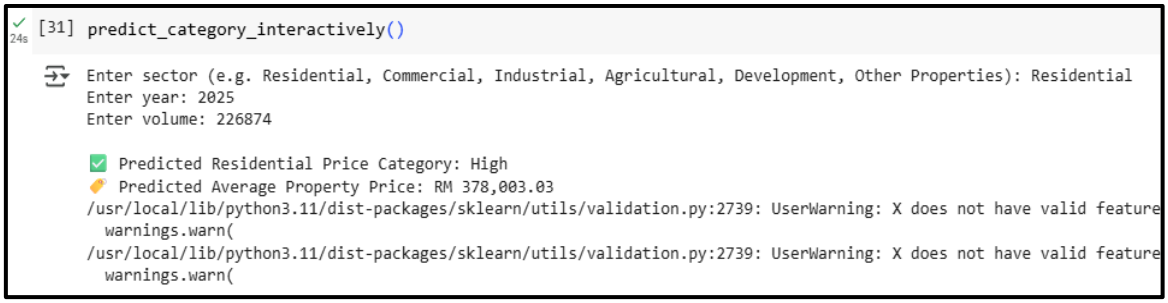


Figure 4: Prediction Output using Random Forest model

To satisfy Objective 3, which is to predict property prices using the best model chosen from Objective 2, the Random Forest model was used to predict property prices based on sector, year, and transaction volume. For the Residential sector with a transaction volume of 226,874 in the year 2025, the model classified the price category as High with a predicted average property price of RM378,003.03, as shown in Figure 4. This result reflects a continued upward trend in residential prices and demonstrates the model's ability to handle future data accurately with an accuracy rate of 0.80.

CONCLUSION

This study successfully applied machine learning and data visualisation to predict property prices in Malaysia. The first objective was achieved through an interactive Tableau dashboard that visualised transaction volumes and values across sectors and years, highlighting the residential sector's dominance. For the second objective, Decision Tree and Random Forest models were developed, with Random Forest demonstrating superior performance in accuracy, precision, recall, and F1-score. The third objective was fulfilled by using Random Forest, the best-performing model, to predict property price categories with higher accuracy, confirming its reliability for stakeholders. By integrating predictive modelling with interactive visualization and proposing enhancements such as richer datasets and user engagement, the study provides valuable insights into market trends and supports the development of a smart property platform that empowers data-driven decision-making in Malaysia's property market.

ACKNOWLEDGEMENT

I would like to extend my sincere appreciation to my supervisor, Dr. Siti Zanariah Satari for her invaluable guidance, encouragement, and constructive feedback. My gratitude also goes to Universiti Malaysia Pahang Al-Sultan Abdullah and Centre for Mathematical Sciences for providing the facilities and supportive environment that enabled the completion of this work. Finally, I am profoundly grateful to all individuals who have directly and indirectly contributed to this endeavour.