International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

MACHINE LEARNING MODELS FOR REAL ESTATE PRICE FORECASTING

A.Nishanth[1], Md. Ghouse [2], Hemanth [3], Dr. N. Krishnaiah [4]
[1,2,3] UG Scholar, Dept of IT, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4]Assistant Professor, Dept of IT, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
nishanthanupuram@gmail.com

Abstract:

.

In the ever-evolving landscape of real estate, house prices play a pivotal role in shaping economic stability and individual decision-making. The valuation process often encounters numerous challenges, necessitating a data- driven approach to ensure accuracy and reliability. To address these complexities, our model integrates meticulous feature engineering, which involves advanced data cleansing, transformation, and feature selection. By leveraging sophisticated machine learning techniques, we aim to capture the underlying trends influencing property prices. The optimization phase incorporates hyperparameter tuning and cross-validation, ensuring that our predictive framework effectively generalizes across diverse datasets. This comprehensive methodology allows us to extract meaningful insights and refine valuation dynamics for more informed decision-making. Our approach is anchored in the application of supervised learning algorithms, such as linear regression and K-fold cross-validation, which are adept at identifying intricate relationships within the dataset. These techniques enable the model to discern latent patterns that influence property valuation, enhancing its predictive capabilities. To measure the performance and reliability of our model, we employ evaluation metrics such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and R- squared. These metrics provide a robust assessment of the model's accuracy, ensuring that our predictions align closely with real-world trends. Through this advanced machine learning framework, our research presents a transformative approach to real estate valuation, paving the way for more precise and data-driven pricing strategies.

Keywords: Machine learning, MSE, RMSE, R-squared, K-fold, Real estate, hyperparameter tuning, feature selection.

1. INTRODUCTION

Determining home values in the ever-changing real estate market is a complex task with far-reaching consequences for individuals, communities, and broader economic systems. Property valuation is influenced by numerous factors, including location, economic conditions, and market demand. Traditional valuation techniques, while useful, often fail to capture the nuanced and dynamic nature of real estate pricing The intrinsic complexity of this process underscores the need for a data-driven approach that
incorporates modern technological advancements to provide a comprehensive understanding of the various factors influencing home prices.

Accurate housing price prediction necessitates a shift from traditional methodologies to more sophisticated, data-centric models. The core challenge lies in predicting home prices while accounting for a wide array of factors, such as market fluctuations, economic trends, and property characteristics. To address this challenge, our model employs an extensive feature engineering process that includes advanced data cleansing and transformation techniques. This ensures that the data fed into our model is well-structured, relevant, and capable of capturing meaningful trends.

2. LITERATURE SURVEY

Abut, Adıgüzel,et.al (2023) [1] "A New Hybrid Approach for Real Estate Price Prediction Using Outlier Detection, Feature Selection, and Clustering Techniques," 2023 Economic risk is a probability that measures the possible alterations, as well as the uncertainty, generated by multiple internal or external factors. Sometimes it could cause the impossibility of guaranteeing the level of compliance with organizational goals and objectives, which is why for their administration they are frequently divided into multiple categories according to their consequences and impact. Global indicators are dynamic and sometimes the correlation is uncertain because they depend largely on a combination of economic, social, and environmental factors. Thus, our proposal consists of a model for prediction and classification of multivariate risk factors such as birth rate and population growth, among others, using multiple neural networks and General Type-2 fuzzy systems. The contribution is the proposal to integrate multiple variables of several time series using both supervised and unsupervised neural networks, and a generalized Type-2 fuzzy integration. Results show the advantages of utilizing the method for the fuzzy integration of multiple time series attributes, with which the user can then prevent future threats from the global environment that impact the scheduled compliance process.

M. S. Bennet Praba et.al. (2023) [2] "Real Time Automation on Real Estate using API," Real estate can be confusing, unclear, disoriented and many a times the price or listings are put up randomly what the
seller decides. Such a system makes it difficult for buyers to make reliable decisions also for a seller to determine what price must he put up the listing or what their property is worth.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

to public transport, cities, neighborhood and cultural aspects, availability, furnish status, etc. A model has to be deployed to provide accurate real estate decisions that are based on a variety of features and tags related to the property. To estimate such a value real time, we need a real time data source to help a community or individual determine the actual deserving value for their property they wish to buy or sell. This system looks to deploy well designed models that can adjust to variations in data due to geographical, economic and political differences by modelling real time data using an API

S. Prongnuch, et.al. (2023) [3] "Outcome-based Learning in Online STEM Activities for Robot and Real Estate Management Camp,". This paper presents the outcome-based learning STEM online activities for the robot & real estate and facility management camp (R^2 Camp) in order to develop the soft skills within knowledge management. The main objective is to provide knowledge on the real estate & facility management, the robotics engineering and how to apply robotics application to the real estate management, which can inspire young people to study in the robotics engineering and real estate & facility management. There are two sections of STEM online activities as: 1) Fundamental of real estate management and robotics engineering, 2) Robotics application in the real estate management. Pre-test and Post- test about the both fundamental are used for outcome-based assessment. Results show that learners' outcomes can achieve the basic knowledge by 10.4%.

Wandhe, Sehgal, Sumra, et.al. (2022) [4], "Real Estate Prediction System Using ML," In recent years, machine learning has played a significant role in many aspects of our lives, including medical diagnosis, natural speech command, picture detection, product suggestion, spam recognition, and price prediction, etc. The desire to receive a profit on an investment property is a typical justification for home buying. They frequently want to know when and where to buy a property, thus they frequently ask the same questions. Current Real Estate Management System does not provide the prediction of the property price for users. In this project we offer the facility to the users to look for properties. This will provide the facility to view the system as an admin or user. The unique feature of this system is that it can predict the chance of a rise in the price of a property in the coming future. Prediction is done with the help of Machine Learning Algorithms such as Convolutional Neural Network (CNN) and Natural Language Processing (NLP), which helps to generate an output that is either positive (1) or negative (0). The predicted value depends upon various factors of the property such as property area, surrounding area, age of construction, Floors, Available rooms and carpet area. This will help the user to predict the chance of the property price rise in the future, thus making a better deal.

Zheng, Yang, Zhang, Z. Bai, Sun et.al. (2022) [5], "Mass Appraisal of Real Estate Prices Using Improved BP Neural Network with Policy Evaluation,". This paper proposes a new policy index neural network (NN), which uses the back propagation (BP) NN model optimized by genetic algorithm (GA) and particle swarm algorithm and the policy modelling consistency (PMC) index to quantify the relevant real estate policies. This study also trains and validates the big data of primary housing transactions of 33 properties in the main urban area of Weihai City, Shandong Province and evaluates the model performance using mean squared error, mean absolute percentage error, and R^2 indexes. Subsequently, it compares the model performance under the GA and particle swarm algorithm optimization, respectively. The results show

that the BPNN model with policies as independent variables is more accurate in predicting house prices in Weihai City.

Y. Sun and G. Peng et.al. (2022) [6], "Developing Area Real Estate Valuation Based on Linear Regression and KNN Algorithm," The study identifies critical factors influencing companies' operational and sustainability performance utilising fluid power systems. Firstly, the study performs Machine Learning (ML) modelling using variables extracted from survey instruments in the West Balkan region. The dataset comprises 115 companies (38.75% response rate). The survey data consist of 22 predictors, including meta-data and three target variables. The K-Nearest Neighbours algorithm offers the highest predictive accuracy compared to the other seven ML models, including Ridge Regression, Support Vector Regression, and ElasticNet Regression. Next, using a model-agnostic interpretation, we assess feature importance using mean dropout loss. After extracting the most essential features, we test hypotheses to understand individual variables' local and global interpretation of maintenance performance metrics. The findings suggest that Failure Analysis Personnel, data analytics, and the usage of advanced technological solutions significantly impact the availability and sustainability of these systems.

G. Wang et.al. (2022) [7] "Applying Internet of Things Framework in Real Estate Business with Enterprise Architecture Approach," The real estate market in Indonesia is currently undergoing a downturn due to a variety of issues, including potential purchasers being unable to find houses that meet their criteria, a lack of time to do home assessments due to the pandemic, and the difficulties of transporting furnishings to new residences. Because development is one of the sectors that affect national economic growth, these issues must be addressed. The purpose of this research is to create an Enterprise Architecture design for the real estate market that incorporates Internet of Things (IoT) technology. The methodology for this research includes observation, data gathering and interviews which addressed the real estate industry's difficulties and demands, followed by designing the proposed business application and Enterprise Architecture Framework. This research found an Enterprise Architecture (EA) design for the real estate business that incorporates IoT technology.

Balasooriya et al., (2022) [8]"Location Intelligence Based Smart E-Commerce Platform for Residential Real-Estate Industry," Making the decision to purchase or invest in real estate can be a very crucial process due to its high financial risk. The purchasing decision of residential real estate properties can be even more decisive because, apart from the financial risk, the choice of a property can have a great impact on the future lifestyle of the buyer. When considering residential real estate, one major factor to be considered is the property location. This research sought to determine the applicability of modern technologies such as location intelligence and machine learning in the development of an e-commerce system that may assist users in making optimal residential real estate location decisions. Third-party web APIs were used to obtain location data, and as a result of the study, methodologies were defined to convert location data into meaningful insights by using statistical methods like weighted sum and the analytical hierarchy process. The functionalities in the proposed system have been designed, considering the roles of both buyers and sellers in the real estate business. In the proposed system, the location quality index framework provides overall insights on the location, the personalized insights and alternative locations are generated via the personal preference-based suitability analyser and the price prediction system provides the current and future price fluctuations. The usefulness of image processing technologies and machine learning for making the sellers' journey easier on a real estate platform has also been assessed.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

3. PROPOSED METHODOLOGY

This proposed methodology focuses on predicting house prices based on key features such as house age, distance from the nearest MRT station, number of convenience stores nearby, latitude, and longitude. The primary objective of this model is to provide accurate and reliable real estate price predictions using machine learning techniques. The model leverages multiple algorithms, including Linear Regression, Support Vector Machine (SVM), and Random Forest, to analyse property data and predict housing prices based on historical trends and spatial attributes.

The system is designed to handle various complexities in housing price prediction, ensuring robustness and efficiency in different real estate market conditions. The methodology involves multiple steps, from data collection to model evaluation, ensuring high predictive accuracy.

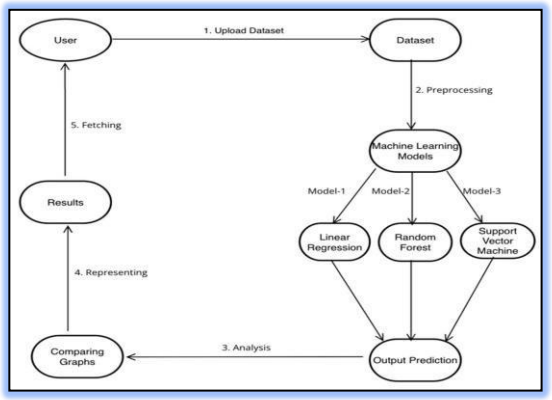


Figure 1: Proposed Machine learning prediction system.

Linear Regression:

The Linear Regression algorithm is a fundamental predictive modelling technique used in this proposed system for real estate price prediction. It establishes a relationship between the dependent variable (house price) and multiple independent variables, including house age, distance to the nearest MRT station, number of convenience stores nearby, latitude, and longitude. The algorithm operates by fitting a linear equation of the form:

$$[Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon]$$

where Y represents the predicted house price, X1, X2, X3, X4 and X5 are the input features, β0\beta_0β0 is the intercept, β1, β2, β3, β4, and β5 are the feature coefficients, and ϵ is the error term. The model learns these coefficients by minimizing the Mean Squared Error (MSE) between actual and predicted values through an optimization process called Ordinary Least Squares (OLS). Linear Regression is preferred due to its simplicity, interpretability, and ability to provide insight into how each feature influences house prices. Despite its limitations in handling non-linearity and complex relationships, it serves as a strong baseline model for understanding the key factors affecting real estate pricing.

The proposed methodology encompasses several key components to ensure accurate real estate price prediction. Data Collection involves gathering a dataset containing crucial features such as house age, distance to the nearest MRT station, number of convenience stores nearby, latitude, and longitude, as these factors significantly influence property valuation. Data Preprocessing is performed to handle missing values, normalize data, and apply feature scaling, ensuring improved model performance and stability. Feature Selection and Engineering focuses on identifying the most relevant features through correlation analysis and domain knowledge to maximize predictive accuracy and enhance model efficiency. Model Selection involves training and evaluating three machine learning models—Linear Regression, Support Vector Machine , and Random Forest—to determine the best-performing model based on both interpretability and predictive power.

The dataset is then divided into training and testing sets, utilizing K-Fold Cross-Validation to ensure an unbiased evaluation by reducing overfitting and improving generalization. Model Evaluation is carried out using performance metrics such as Mean Squared Error (MSE), R-Squared ($\mathbb{R}^2$), and Root Mean Squared Error (RMSE) to assess prediction accuracy, variance, and generalization capability. To further enhance performance, Hyperparameter Tuning is applied using techniques like Grid Search or Random Search to optimize model parameters such as the number of trees in Random Forest or the kernel type in SVM. Finally, the Prediction and Deployment phase involves using the optimized model to predict house prices, which can be integrated into a web-based application or API service, allowing users to input property details and receive price estimations in real-time. This model can be further refined with additional real-world data, improving its accuracy and adaptability to changing market conditions.

Applications

The housing price prediction model has several practical applications across different industries. In Real Estate Market Analysis, it helps buyers, sellers, and investors make informed decisions by analysing historical price trends and predicting future values. Urban Planning and Development can leverage the model to assess how factors like transportation facilities, shopping centres, and schools impact property values, aiding in strategic city planning. Banking and Mortgage Lending institutions use the model to evaluate property values, assess loan risks, and determine appropriate mortgage amounts for borrowers. Additionally, Property Investment Strategies benefit from predictive insights, allowing investors to identify profitable locations, optimize investment portfolios, and minimize risks.

Advantages

The proposed model offers multiple advantages that enhance its reliability and efficiency in real estate price prediction. Accurate Predictions are ensured through the integration of multiple machine learning algorithms, improving the precision of price forecasting. The model's Scalability enables it to process large datasets efficiently, making it applicable across diverse housing markets. Feature Importance Analysis provides insights into the key factors influencing property prices, helping stakeholders make data-driven decisions. The model's User-Friendly Interface, when deployed as a web application, allows users to input property details and instantly obtain estimated prices. Additionally, Data-Driven Decision Making enables buyers, sellers, and investors to rely on historical trends and predictive analytics, enhancing confidence in real estate transactions.

4. EXPERIMENTAL ANALYSIS

The experimental analysis gives the performance of the project in builder's ecosystem. The dataset used for training and evaluation contains five key features: House Age, Distance to the Nearest MRT Station, Number of Convenience Stores Nearby, Latitude, and Longitude. The model's performance is assessed using various evaluation metrics and visualizations.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

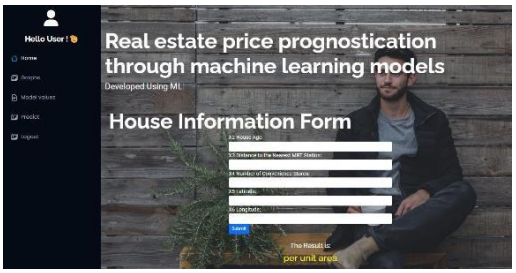


Fig 2 Prediction page

Figure 1 shows the prediction page of the web-based application, where users can input property details such as house age, distance to the nearest MRT station, number of convenience stores nearby, latitude, and longitude. Once the details are submitted, the trained machine learning model processes the input to generate a predicted house price.

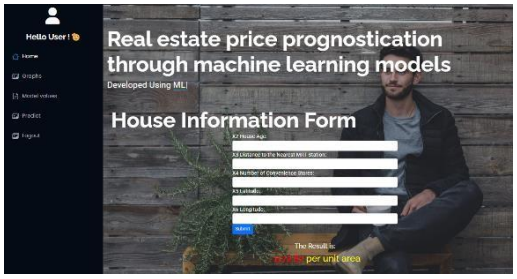


Fig 3 Output page

Figure 2 displays the output of the prediction page, showcasing the predicted price based on the provided property attributes. This feature enables users, real estate investors, and buyers to quickly estimate property values based on market trends and key influencing factors.

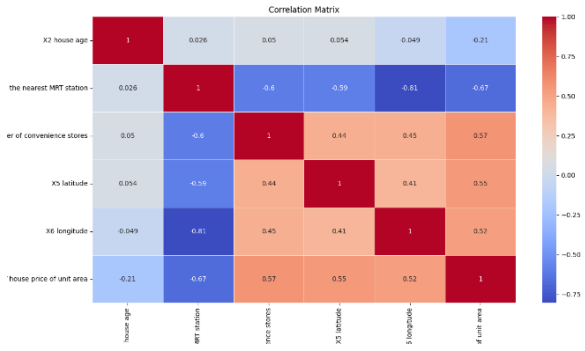


Fig 4 Corelation Matrix

Figure 4, This heatmap illustrates the correlation between input features and house prices. It provides insights into how strongly each factor affects property values. The matrix indicates that distance to the MRT station has the highest negative correlation with house prices, reinforcing the significance of location in real estate valuation.

Fig 5 Scatter plot

This scatter plot compares the actual house prices with the predicted values. A well-performing model should exhibit a linear trend where predicted values closely align with actual prices. The visualization helps evaluate the consistency and accuracy of the trained model.

5. CONCLUSION

The study presented here investigated on the use of Support Vector Machine, Random Forest, and Linear Regression models for real estate price prediction. The base was laid via meticulous preparation, which included data cleaning and outlier reduction. Evaluation metrics demonstrated the balanced performance of Linear Regression, while visual analysis demonstrated the robust predictive strength of the Random Forest model. The results highlight the fine balance that must be struck between interpretability and accuracy. Linear regression was found to be a reasonably balanced option, while Random Forest demonstrated exceptionally high accuracy. Although reasonable, Support Vector Machine displayed more mistakes. This study provides information to help with decision-making in the ever- changing housing market. Subsequent improvements can concentrate on exploring features through feature engineering, improving the model, and using cutting-edge methods to improve prediction accuracy.

REFERENCES

[1] Z. F. Abut, H. Ş. Arlı, M. F. Akay and Y. Adıgüzel, "A New Hybrid Approach for Real Estate Price Prediction Using Outlier Detection, Feature Selection, and Clustering Techniques," 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, 2023, pp. 1-6, doi: 10.1109/UBMK59864.2023.10286673.

[2] M. Arivukarasi, A. Manju, R. Kaladevi, S. Hariharan, M. Mahasree and33, 2023, pp. 1-4, doi: 10.1109/iSTEM-Ed55321.2022.9920828.

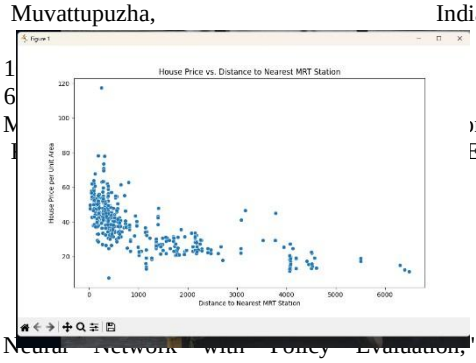
[3] A. Wandhe, L. Sehgal, H. Sumra, A. Choudhary and M. Dhone, "Real Estate Prediction System Using ML," 2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP), Nagpur, India, 2023, pp. 1-4, doi: 10.1109/ICETET- SIP58143.2023.10151561.

[4] A. Peter, A. A. Kumar, A. Rajeev, B. Baiju and V. S. Chooralil, "Real Estate Management System using Blockchain," 2023 International Conference on Innovations in Engineering and Technology (ICIET),

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

Muvattupuzha, India, 2023,



[5] M ore and A. Estate using

[6] Y. Zheng, B. Yang, R. Zhang, Z. Bai and Y. Sun, "Mass Appraisal of Real Estate Prices Using Improved BP API," 2023 First International Conference on Electrical, Electronics, Information and CommunicationTechnologies (ICEEICT), Trichy, India, 2022, pp. 1-5, doi: 10.1109/ICEEICT53079.2022.9768428.

[7] N ' 2023 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Dalian, China, 2023, pp. 1036-1041, doi: 10.1109/TOCS56154.2023.10015915.

[8] Y. Sun and G. Peng, "Developing Area Real Estate Valuation Based on Linear Regression and KNN Algorithm," 2023 6th Annual International Conference on Data Science and Business Analytics (ICDSBA), Changsha, China,2022, pp.3842, doi:10.1109/ICDSBA57203.2023.00014

[9] S. Thokala, R. Jakkani, M. Alli and H. Shanmugasundaram, "Accident Prevention System using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-6, doi: 10.1109/ICCCI56745.2023.10128202.

[10]H. A. V. P. U. Hapuarachchi, M. D. Manoratne, K. G. B. K. Gamlath, V. S. G. G, D. Sriyaratna and N. H. P. R. Supunya, "Realty Scout Smart System for Real Estate Analysis & Forecasting with Interactive User Interface," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-6, doi: 10.1109/I2CT54291.2022.9825335.

[11] W. Coleman, B. Johann, N. Pasternak, J. Vellayan, N. Foutz and H. Shakeri, "Using Machine Learning to Evaluate Real Estate Prices Using Location Big Data," 2022 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA,2022,pp.168-172,\doi: 10.1109/SIEDS55548.2022.9799393.

[12]Y. Zhao, X. Shen, X. Xu and Y. Xu, "Application of BP neural network in real estate batch assessment," 2022 41st Chinese Control Conference (CCC), Hefei, China, 2022, pp. 7018-7023, doi: 10.23919/CCC55666.2022.9901571.

[13]Y. Zhao, G. Chetty and D. Tran, "Deep Learning for Real Estate Trading," 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 2022, pp. 1-7, doi: 10.1109/CSDE56538.2022.10089222.

[14]S. Wang, J. Zhu, Y. Yin, D. Wang, T. C. Edwin Cheng and Y. Wang, "Interpretable Multi-Modal Stacking-Based Ensemble Learning Method for Real Estate Appraisal," in IEEE Transactions on Multimedia, vol. 25, pp. 315-328, 2022, doi: 10.1109/TMM.2021.3126153.

[15]T. Zhu, "The Impact of Non-immersive Virtual Reality Technologies on Consumers' Behaviors in real estate: A Website's Perspective," 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, Singapore, 2022, pp. 13-20, doi: 10.1109/ISMAR-Adjunct57072.2022.00013.

[16]L. Kong et al., "When permissioned blockchain meets IoT oracles: An on-chain quality assurance system for off-shore modular construction manufacture," 2022 IEEE 1st Global Emerging Technology Blockchain Forum: Blockchain & Beyond (iGETblockchain), Irvine, CA, USA, 2022, pp. 1-6, doi: 10.1109/iGETblockchain56591.2022.10087164.

[17]R. Henker, D. Atzberger, W. Scheibel and J. Döllner, "Real Estate Tokenization in Germany: Market Analysis and Concept of a Regulatory and Technical Solution," 2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Dubai, United Arab Emirates, 2023, pp. 1-5, doi: 10.1109/ICBC56567.2022.10174954.

[18] T. Balasooriya et al., "Location Intelligence Based Smart E-Commerce Platform for Residential Real-Estate Industry," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 867-873, doi: 10.1109/ICOSEC54921.2022.9952023.

[19]M. Selim, M. R. Rabbani and A. Bashar, "Qard Hasan Based Cooperative Model for Home Financing and Its Effects in Home Ownership and Real Estate Development," 2022 International Conference on Sustainable Islamic Business and Finance (SIBF), Sakhir, Bahrain, 2022, pp. 48-52, doi: 10.1109/SIBF56821.2022.994002.

[20]G. Wang, J. S. Suroso, D. Sanusi, J. A. Tanuwijaya and T. F. I. Theodora, "Applying Internet of Things Framework in Real Estate Business with Enterprise Architecture Approach," 2022 International Conference on Information Management and Technology (ICIMTech), Semarang, Indonesia, 2022, pp. 219-224, doi: 10.1109/ICIMTech55957.2022.9915151.