Asian Journal of Engineering and Applied Technology
ISSN: 2249-068X (P)
Vol.13 No.1, 2024, pp.37-43

www.crijournals.org
DOI: https://doi.org/10.70112/ajeat-2024.13.1.4240

Enhancing Rental Cost Predictions for Student Housing in Lokoja: A Comparative Analysis of Machine Learning Models

Simon Ojima Abuh, Fati Oiza Ochepa and Malik Adeiza Rufai
Department of Computer Science, Federal University Lokoja, Nigeria
E-mail: abuhsimonoj@gmail.com, fati.salami@fulokoja.edu.ng, rufai.malik@fulokoja.edu.ng
(Received 7 March 2024; Revised 14 April 2024; Accepted 22 April 2024; Available online 28 April 2024)

***Abstract* - Accurately determining rental costs for properties in specific locations is crucial for influencing market dynamics and housing decisions for both homeowners and prospective tenants. This study presents a data-driven predictive analytics model designed to forecast house rents for students in Lokoja, North-Central Nigeria, using machine learning techniques. The research involved meticulous data collection from students residing within the three major student-dominated localities of Adankolo, Felele, and Crusher in Lokoja. Data preprocessing and feature selection were undertaken to ensure the quality and relevance of the dataset. The dataset comprises eleven predictors and 300 observations. Eventually, seven predictors and 265 observations were used for modeling, with a split of 80% for model training and 20% for testing. Three machine learning algorithms - Random Forest Regression, Linear Regression, and Decision Tree Regression—were evaluated for their predictive accuracy. The Random Forest Regression model emerged as the most accurate, achieving an R2R^2R2 value of 95.2%. Correlation analysis confirmed a strong positive relationship between the selected features (e.g., number of rooms, furniture) and rental prices. A user-friendly interface was developed to facilitate rent predictions based on user inputs. The findings underscore the model's robustness and its potential for real-world application in the real estate market. Suggestions for future work include incorporating additional features, expanding the geographic scope, and exploring advanced machine learning techniques to further enhance model accuracy and applicability.**
***Keywords:* Predictive Analytics, Machine Learning, Rental Costs, Random Forest Regression, Student Housing**

I. INTRODUCTION

The real estate sector plays a crucial role in driving economic growth, impacting both the overall economy and individual businesses. It significantly influences employment, income distribution, and the economic well-being of regions and their inhabitants. Housing rents undoubtedly affect living standards, as they often constitute a substantial share of household expenses. Accurate rental predictions can influence investment choices, housing affordability, and urban development [1]. Housing rents, being an essential component of the quality of life in various regions, impact both existing residents and prospective relocators. Rent levels are a major determinant of individuals' and families' ability to find suitable homes within their budget, which in turn affects their quality of life, financial independence, and general well-being.

Studies worldwide have shown that rental housing is crucial as it meets one of the basic human needs. It is often the first entry point for prospective relocators in a city. Until existing residents and prospective relocators can save enough to invest in ownership housing, rental housing provides essential shelter, thereby reducing the accommodation burden on inhabitants. The demand for housing is gradually increasing with population growth. People require homes equipped with their preferred amenities in their desired neighborhoods [2].

Accurately predicting house rent prices is a critical issue of interest to researchers, real estate agents, homeowners, and prospective renters. Precise rent price prediction can provide valuable insights for stakeholders seeking affordable housing, as predictive analytics use historical data and advanced algorithms to forecast future trends and outcomes [1]. However, predicting rental prices accurately is challenging due to the complex interplay of various factors influencing house rental prices. These factors include location, house features, availability of amenities, economic conditions, and overall supply and demand in the rental market. Machine learning models have been applied to diverse rental market datasets, real estate databases, and government records. Several studies have explored machine learning models for predicting house rent prices. For instance, a study by O. Oshodi *et al.,* [2] applied a neural network model to estimate rental prices of residential properties using a dataset from Cape Town, South Africa, incorporating fourteen property attributes. The results showed that the neural network model achieved a prediction accuracy of 78.95%, outperforming traditional models in terms of accuracy.

With increasing urbanization, more people are moving to cities, driving a continuous rise in housing demand. One of the major challenges for those seeking rental housing is the lack of information on appropriate rental prices [3]. Predicting house rents is complex due to the myriad factors influencing rental prices. Existing methods of estimating house rents are often based on assumptions or rough estimates, such as comparing rent prices for similar properties in the same area. These methods rely on manual analysis of property features, which is time-consuming and prone to human error.

This study aims to develop a machine learning-based tool for accurately predicting house rent prices, thereby enhancing decision-making processes across the real estate sector. The findings of this study have the potential to significantly contribute to economic development, housing affordability, and market efficiency, benefiting a broad range of real estate stakeholders.

II. REVIEW OF LITERATURE

Efforts and research have been directed toward improving the quality and accuracy of house rent prediction models using various machine learning algorithms and techniques. This section outlines several research works and the techniques adopted in developing accurate prediction models.

A blend of machine learning and deep learning techniques was developed by B. Mallikarjuna *et al.,* [4] to predict house prices using regression algorithms. The dataset, sourced from Kaggle, consisted of fourteen house features. The results demonstrated high accuracy in estimating house values based on the documented features. Additionally, the researchers developed a web-based smart home application from the model to assist users in predicting house prices.

The study conducted by K. Shah *et al.,* [5] utilized regression techniques to build a predictive model for house rent prices. The dataset included thirteen features impacting house rent prices. Techniques such as Gradient Boosting Regressor, Random Forest Regressor, Multilayer Perceptron Regressor, and AdaBoost Regressor were applied. The model's performance was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

A study by Y. Ming *et al.,* [6], based on an online housing platform, utilized a dataset from Chengdu housing rentals, comprising 33,224 items with features such as housing area, structure, mode of rent collection, and means of transportation. Models developed using Random Forest Regressor, XGBoost, and LightGBM were trained and tested. Data visualization revealed that occupants favored smaller area apartments. Accuracy evaluation indicated that XGBoost performed the best, with an accuracy of 85%.

Another study by Y. Kang *et al.,* [7] proposed a data-fusion framework to examine house price prediction by combining multiple data sources. The dataset, which included seven thousand houses in the Greater Boston Area, incorporated house structural attributes, photos, locational amenities, street view images, transportation accessibility, visitor patterns, and socioeconomic attributes of neighborhoods. The Gradient Boosting Machine algorithm was used to build the model, providing insights into how multi-source geo-data can enhance predictive modeling of real estate price trends. Furthermore, H. Kim *et al.,* [8] employed the Long Short-Term Memory (LSTM) technique to predict house prices in Busan, South Korea, based on 547,740 apartment transaction records. Features and proximity measures to public rental housing (PRH) were derived. Evaluation across four geographic boundaries indicated that the model provided accurate and dependable predictions, highlighting the significant impact of proximity to PRH on housing prices. A study by T. Mohd *et al.,* [9] proposed a novel technique to identify the most efficient model for predicting office rent and the most influential features. The study utilized Decision Tree, Random Forest, and Support Vector Machine algorithms. The Decision Tree algorithm outperformed the other models. The results indicated that available amenities and in-house services were the two major features impacting office rent prices.

Additionally, K. Lee *et al.,* [10] developed a machine learning-based model for rental house prediction using the Random Forest Regression model on two datasets: AirBnB Seoul and Tokyo, and a global dataset. The model's performance was evaluated using MAE, MSE, and Root Mean Absolute Error (RMAE). The evaluation demonstrated that the system allowed users to input house rent information and generate highly accurate predictive results.

The study conducted by Y. Zhang *et al.,* [11] addressed gaps in existing house rent prediction approaches based on hedonic price modeling. The researchers built a model using the Random Forest algorithm with a comprehensive dataset from housing transactions in Toronto. The model achieved an accuracy of 85%, revealing that socioeconomic factors, house features, and building accessibility significantly impact house rent prices. The model was recommended as a valuable tool for decision-makers and real estate planners. A predictive house price model based on hood structure and selected residential rental pricing for houses in Petaling Jaya, Selangor, was developed by R. K. Ayyasamy *et al.,* [12]. The study employed several machine learning algorithms, including a geographic data handler for map-based visualization of local amenities, a rental analytics unit for current price trends, and a price-based predictive module. The XGBoost algorithm demonstrated the best prediction accuracy.

Overall, various approaches can be employed to build rental prediction models. The choice of approach depends on the specific data available and the intended application of the model.

III. METHODOLOGY

This section outlines the technical steps and procedures involved in developing and deploying the proposed house rent prediction model. The process began with the researchers curating the dataset by directly collecting data from residents in the three major student residential areas in Lokoja, Nigeria. The data, initially unstructured, was subjected to cleaning and preprocessing. This section also discusses the input and output designs, which define the modules and interfaces, as well as the architectural and logical designs of the proposed model.

A. Architectural Design of the Proposed Model

The architectural design facilitates understanding, analysis, communication, and decision-making by providing an overview of how the main components will be constructed. It offers a high-level view of the structure and interconnections among the various components involved in building the model. Fig. 1 illustrates the architectural design of the proposed model.

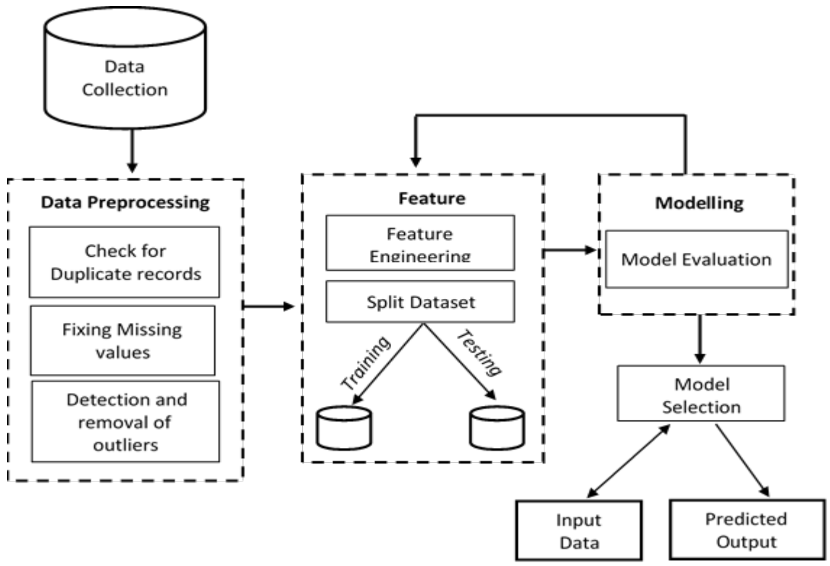

Fig. 1 Architectural design of house rent prediction model

B. Architectural Design of the Proposed Model Logical System Design

Logical design involves the conceptual representation of interactions between the functional components of a system. It provides an abstract representation of a system's input, output, database architecture, and data flow, which are not directly visible to the user. In the context of this study, and as depicted in Fig. 2, the logical design includes identifying the key components of the system: input data (e.g., location of the apartment, nature of the apartment, number of rooms), data processing algorithms (e.g., machine learning models), and output predictions (e.g., estimated rent prices).

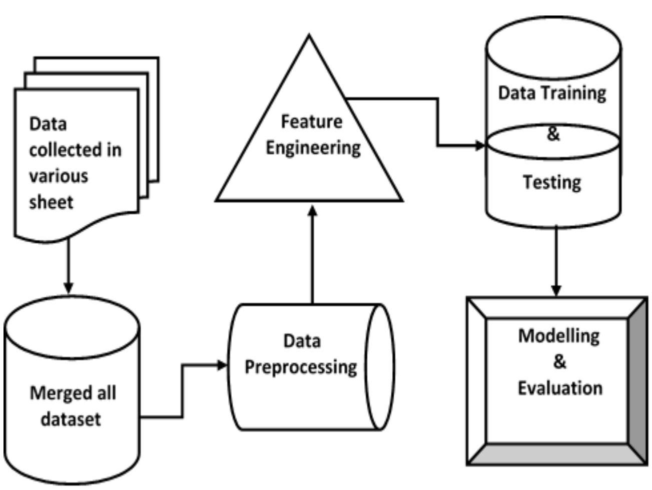

Fig. 2 Logical design diagram of house rent prediction model

C. Use Case Diagram

The use case diagram visually represents the interaction between the users (actors) and the house rent prediction model, illustrating the functionality provided by the system and the actors that interact with it. The use case diagram in Fig. 3 shows the users (landlord, tenant) as the primary actors who interact with the Rent Prediction Model to estimate the rent of a house or apartment. The model takes input from the user and provides output based on the selected house features.

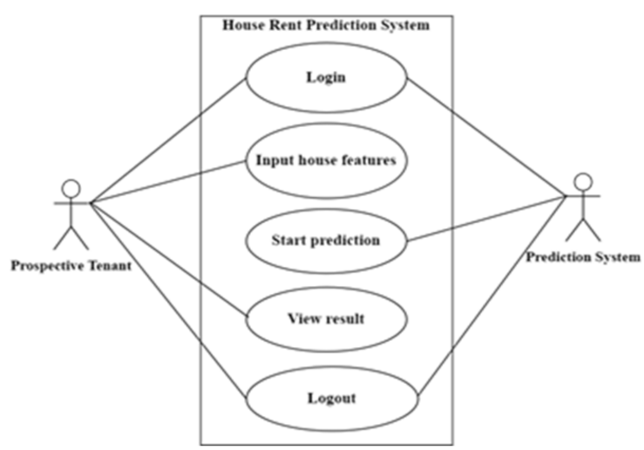

Fig. 3 Use case diagram of house rent prediction system

This section details the methodology employed in developing the house rent prediction model. It begins with the architectural design, offering a comprehensive overview of the model's structure and the interaction between its main components. The logical system design follows, addressing the abstract representation of system inputs, outputs, database architecture, and data flow. The use case diagram visually illustrates the interaction between users and the prediction model, highlighting the system's functionality. Together, these sections lay a solid foundation for understanding the technical steps and procedures required to construct an accurate, reliable, and effective house rent prediction model.

IV. RESULTS AND DISCUSSION

In this section, the results obtained from translating user requirements into a tangible prediction model are presented. The dataset used to test the model's accuracy was manually curated by the researchers. The results were evaluated to determine the accuracy of the developed model.

A. The Dataset

The dataset curated for this research project included house features such as location, nature of the apartment, and rent price. The data was collected primarily from three major residential areas - Adankolo, Felele, and Crusher - in the Lokoja metropolis, specifically around the Federal University Lokoja, Nigeria. The collected dataset comprised eleven predictors and 300 observations. Fig. 4 depicts the visualization of the dataset instances.

```
In [13]: dataset
Out[13]:
     Locality Type of House  ... Numbers of bathroom Rent Price
0    Adankolo     Apartment  ...                   1      60000
1      Felele     Apartment  ...                   1     120000
2    Adankolo     Apartment  ...                   1     130000
4    Adankolo     Apartment  ...                   1     130000
5    Adankolo     Apartment  ...                   1      60000
..        ...           ...  ...                 ...        ...
261    Felele     Apartment  ...                   1     135000
262    Felele         Flats  ...                   1     150000
263    Felele     Apartment  ...                   1     140000
264  Adankolo         Flats  ...                   1     140000
265  Adankolo         Flats  ...                   1     130000

[265 rows x 8 columns]
```

Fig. 4 Dataset visualization

B. Data Pre-Processing and Visualization

Data cleaning and pre-processing tasks were undertaken using the Pandas library. This involved the removal of irrelevant attributes and less influential house predictors or factors such as Rent Status, Timestamp, Username, Town, Occupation, and Possession. Consequently, there are no missing values in the dataset. Eventually, eight predictors and 265 observations were used for modeling. The dataset was split into 80% for training and 20% for testing the model. Fig. 5 shows the steps taken to check for missing values in the dataset and provides a description of the data.

```
In [14]: print(dataset.isnull().sum())
Locality                      0
Type of House                 0
Nature of Apartment/Flats     0
Available Utility             0
Furnishing Status             0
Numbers of bedroom            0
Numbers of bathroom           0
Rent Price                    0
dtype: int64
```

```
In [12]: dataset.describe()
Out[12]:
       Numbers of bedroom  Numbers of bathroom
count               265.0                265.0
mean                  1.0                  1.0
std                   0.0                  0.0
min                   1.0                  1.0
25%                   1.0                  1.0
50%                   1.0                  1.0
75%                   1.0                  1.0
max                   1.0                  1.0
```

Fig. 5 Checking for missing value and dataset description

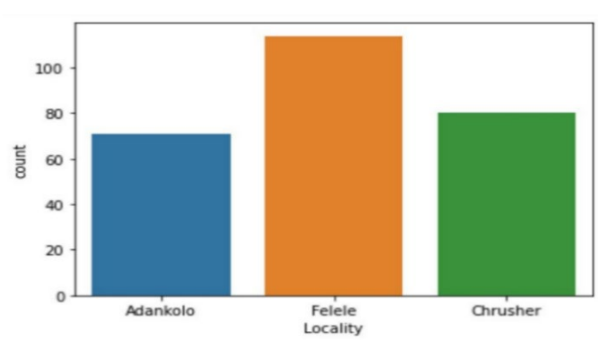

Fig. 6 Total number of houses based on locality

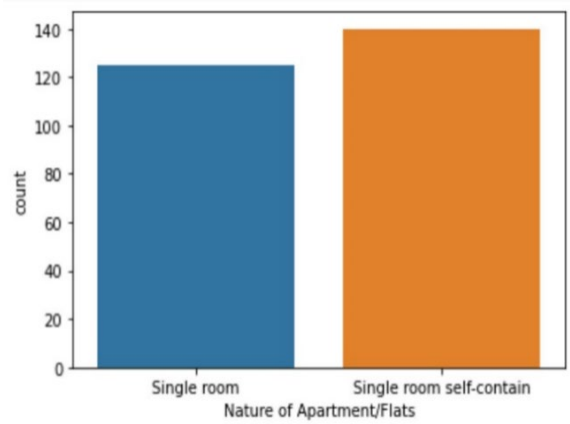

Fig. 7 Total number of single room apartments

Fig. 6 and Fig. 7 illustrate the number of houses by locality and the statistics for total single-room and single-room self-contained apartments from the three localities considered in this study.

C. Presentation of Results

This section presents and interprets the findings and insights derived from the development of the model and its predictions. The results and interpretations are as follows:

1. Model Performance Evaluation: Three machine learning models - Random Forest Regression, Linear Regression, and Decision Tree Regression were used in this study. The evaluation metric employed to assess the performance of the models was the R-squared (R^2) metric, which measures the accuracy and fit of the models. The evaluation results, as shown in Table I, indicate that Random Forest Regression outperformed Linear Regression and Decision Tree Regression, achieving an accuracy of 95%.

TABLE I MODEL PERFORMANCE EVALUATION

Model	Accuracy
Random Forest Regression	95.2%
Decision Tree Regression	94.90 %
Linear Regression	62.30%

2. Correlation: Correlation visually represents data fit by demonstrating a linear relationship between variables. In this research project, the correlation indicated a perfect positive correlation (1), which implies that as one variable increases, the other variable also increases. For instance, as the features of the house increase (e.g., number of rooms, furniture), the house rent also increases. This linear relationship between features is shown in Fig. 8.

Fig. 8 Model Performance evaluation by correlation

3. User Interface: The user interface (UI) is the means through which users interact with and control the software or website to perform specific tasks and access information. For this study, we designed a user-friendly interface using HTML, CSS, and JavaScript. As shown in Fig. 9, it is divided into two sub-interfaces:

i. The Rent Prediction Interface: This interface consists of seven house features or predictors: location, house type, nature of the apartment or flat, number of bedrooms, number of bathrooms, furnishing status, and utility. The user can select one or more options for each feature or predictor. To initiate the house rent prediction, there are two call-to-action buttons: “Click to Predict” and “View Result,” which enable the user to predict house rent based on their chosen preferences and view the predicted price.

ii. Result Display Interface: This interface displays the predicted house rent price based on the selected house features. It includes one call-to-action button, “Predict Again,” which allows the user to initiate a new prediction by reselecting house features or predictors, as illustrated in Fig. 9, to obtain a new predicted price.

Fig. 9 Rent Prediction and Result Display Interfaces

D. Discussion of Results

The results of this study provide significant insights into the performance and efficacy of different machine learning models for predicting house rent prices. To ensure quality and relevance, the data used were meticulously curated and pre-processed, including the removal of irrelevant attributes and addressing any missing values. The initial dataset comprised 11 predictors with 300 observations, collected from three major residential areas in Lokoja, Nigeria. After preprocessing, the dataset was reduced to eight predictors and 265 observations. This cleaned dataset was then split into training and testing sets in an 80:20 ratio. Three machine learning algorithms were employed: Random Forest Regression, Linear Regression, and Decision Tree Regression.

The evaluation metric used was the R-squared (R^2) metric, which measured the accuracy and fit of the models. Random Forest Regression achieved the highest accuracy with an R^2 score of 95.2%, followed closely by Decision Tree Regression with an accuracy of 94.9%. Linear Regression had a significantly lower accuracy of 62.3%. The Random Forest Regression model, having the highest accuracy, was chosen as the most suitable for house rent prediction. Correlation analysis indicated a perfect positive correlation (correlation coefficient of 1), suggesting that an increase in one feature led to an increase in the house rent. This linear relationship validates the model’s predictions and confirms the importance of the selected predictors. The user interface (UI) for the house rent prediction model was designed using HTML, CSS, and JavaScript. It included the rent prediction interface, which allows users to input house features and predict rents, and the rent display interface, which shows the predicted rent price. The findings of this study demonstrate the model’s practical application and its potential for real-world use in predicting house rents based on specific features.

V. CONCLUSION

This research has successfully developed a data-driven predictive analytics model for forecasting house rents in Lokoja, Nigeria. The model utilized machine learning techniques to analyze and predict rental prices based on specific house features. Three machine learning models were employed, with the Random Forest Regression model emerging as the most effective, achieving an R-squared value of 95.2%. This high accuracy underscores the model’s robustness and reliability in predicting house rents. Correlation analysis further validated the model by demonstrating a strong positive relationship between the predictors and rental prices, confirming the importance of features such as the number of rooms and furniture in determining house rent. Additionally, a user-friendly interface was developed to enable users to input house features and receive rent predictions. This practical application enhances the accessibility and usability of the model for various stakeholders, including property managers, real estate agents, and prospective tenants.

VI. SUGGESTION FOR FUTURE WORK

While this study has successfully developed a high-accuracy predictive model for house rents, there are several aspects that could be improved. Future research could explore the inclusion of more diverse and larger datasets from different geographical regions to enhance the model’s generalizability.

Additionally, investigating additional predictive features, such as proximity to amenities, neighbourhood safety, and historical rental trends, could further refine the model’s accuracy. Integrating advanced machine learning techniques, such as deep learning and ensemble methods, may also be beneficial. Furthermore, research could focus on developing real-time data integration capabilities and expanding the user interface to incorporate more interactive and intuitive features, thereby improving the model's practical application and usability for a broader range of users.

REFERENCES

[1] H. Mulchandania, J. Vaswani, R. Vaghela, and R. Patel, "A literature review on house price prediction based on fuzzy logic," *GIT-J. Engineering and Technology*, vol. 14, pp. 162–164, Jul. 2022.

[2] O. Oshodi, W. D. Thwala, T. B. Odubiyi, C. Aigbavboa, and R. B. Abidoye, "Using neural network model to estimate the rental price of residential properties," *J. Financial Management of Property and Construction*, vol. 24, pp. 217–230, Jul. 2019, doi: 10.1108/JFMPC-06-2019-0047.

[3] K. Lamudi, "Majority of Kenyan renters do not know the right price to pay for a house," Nov. 15, 2022. [Online]. Available: https://allafrica.com/stories/201501122440.html

[4] B. Mallikarjuna, S. M. Ram, S. Addanke, and M. Sabharwal, "An improved model for house price/land price prediction using deep learning," in *Handbook of Research on Advances in Data Analytics and Complex Communication Networks*, vol. 12, 2022, doi: 10.4018/978-1-7998-7685-4.ch005.

[5] K. Shah, H. Shah, A. Zantye, and M. Rao, "Prediction of rental prices for apartments in Brazil using regression techniques," in *Proc. 12th International Conference on Computing Communications and Networking Technologies*, India, Nov. 2021, doi: 10.1109/ICCCNT 51525.2021.9579796.

[6] Y. Ming, J. Zhang, J. Qi, T. Liao, M. Wang, and L. Zhang, "Prediction and analysis of Chengdu housing rent based on XG-Boost algorithm," in *Proc. 3rd International Conference on Big Data Technologies*, New York, Sep. 2020, pp. 1–5, doi: 10.1145/3422713.3422720.

[7] Y. Kang, F. Zhang, W. Peng, S. Gao, J. Rao, F. Duarte, and C. Ratti, "Understanding house price appreciation using multi-source big geo-data and machine learning," *Land Use Policy*, vol. 111, Dec. 2021, doi: 10.1016/j.landusepol.2020.104919.

[8] H. Kim, Y. Kwon, and Y. Choi, "Assessing the impact of public rental housing on the housing prices in proximity: based on the regional and local level of price prediction models using long short-term memory (LSTM)," *Sustainability*, vol. 12, Sep. 2020, doi: 10.3390/su12187520.

[9] T. Mohd, M. Harussani, S. Masrom, N. Johari, and L. Alfat, "Office rent prediction based on the influenced features," *Environment-Behaviour Proceedings Journal*, vol. 7, pp. 61–68, Mar. 2022, doi: 10.21834/ebpj.v7i19.3236.

[10] K. Lee, S. N. Njimbouom, and J. Kim, "Development of rent house price prediction service based on machine learning," *Journal of Digital Contents Society*, vol. 23, pp. 2445–2455, Dec. 2022.

[11] Y. Zhang, D. Zhang, and E. J. Miller, "Spatial autoregressive analysis and modeling of housing prices in city of Toronto," *Journal of Urban Planning and Development*, vol. 147, Mar. 2021, doi: 10.1061/(ASCE)UP.1943-5444.0000651.

[12] R. K. Ayyasamy, B. Tahayna, S. Subramaniam, F. Jing Tan, S. Krisnan, and L. N. A. Tahayna, "Design and implementation of residential rental rates forecast model using data mining algorithms," in *Proc. 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, Malaysia, Oct. 2022, doi: 10.1109/AiDAS56890.2022.9918792.