An Integrated Hedonic Pricing and Predictive Modelling Approach: Comparing ML and DL for Dublin's Rental Market

MSc Research Project
Data Analytics

Yihui Zhang
Student ID: x23193824

School of Computing
National College of Ireland

Supervisor: Dr Anu Sahni

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Yihui Zhang
Student ID:	x23193824
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr Anu Sahni
Submission Due Date:	24/04/2025
Project Title:	An Integrated Hedonic Pricing and Predictive Modelling Approach: Comparing ML and DL for Dublin's Rental Market
Word Count:	5638
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Yihui Zhang
Date:	23rd May 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

An Integrated Hedonic Pricing and Predictive Modelling Approach: Comparing ML and DL for Dublin's Rental Market

Yihui Zhang
x23193824

Abstract

Dublin's persistent housing crisis has been a critical issue for a long time. While prior research has primarily focused on the property sales market, the rental market also requires attention, especially as rocketing house prices have forced many residents to rely on rented accommodation. This study bridges this gap by proposing a model integrating hedonic regression with predictive modelling to analyse rental prices in Dublin.

A key challenge was overcome by conducting web scraping to construct a detailed dataset encompassing core features summarised from the literature review. The dataset was further enhanced through spatial analysis by integrating proximity measures to key external amenities.

The findings in this study revealed a significant rental burden issue within Dublin's housing market, with the Rent-To-Income Ratio (RTR) across all household types exceeding the affordability threshold of 30%. Key features that influence rental prices were identified, including property types, building energy ratings (BER), number of bedrooms and bathrooms, and accessibility to neighbourhood and location facilities and amenities. Among Random Forest, SVR, XGBoost, LightGBM, GBR, and RNNs models, LightGBM was found to achieve the optimised predictive accuracy, with an R^2 of 0.79, MSE of 106189.96, MAE of 234.97, $RMSE$ of 325.87, $\%RMSE$ of 13.32%, and $\%MAE$ of 34.61%

1 Introduction

As the capital of the Republic of Ireland, Dublin is the country's financial and cultural centre and is known for its heritage, prosperous tech industry, diverse population, and expensive housing prices. Dublin has a population of approximately 1,534,900 [1], accounting for 28.5% of the total population of Ireland. Of this, nearly 950,000 [2] individuals are employed, representing 36% of the national workforce. However, according to the latest report published by daft.ie[3], the average number of available rental properties in Dublin is fewer than 1,200. This significant imbalance has made rental availability and affordability a lasting challenge in the city. Fuelled by the global financial crisis of 2008-2009

[1]https://www.cso.ie/en/releasesandpublications/ep/p-pme/populationandmigrationestimatesapril2024/keyfindings/
[2]https://www.dublinchamber.ie/About-Us/Economic-Profile-of-Dublin
[3]https://ww1.daft.ie/report/ronan-lyons-2024q4-daftrentalprice?d_rd=1

Boeing and Waddell (2017), the housing crisis in Ireland has intensified. The rising unemployment rate has made home ownership more unaffordable, making renting the only option for many people Lima Holanda (2018).

The Irish government have undertaken several strategies to deal with this issue, including the rental caps strategy and tenant financial support scheme Lima Holanda (2018). However, these efforts achieved little effect. In fact, frequent policy changes have resulted in unstable government regulation and long-term reliance on temporary solutions. This instability was believed to have discouraged long-term investment in social housing construction and failed to fully benefit tenants or address the essential causes of the housing crisis.

The previous study of the Dublin housing market was primarily focused on the purchasing market and used the hedonic regression model to estimate prices Hurley and Sweeney (2024). However, the use of machine learning and deep learning models for price prediction has become increasingly prevalent. A study Hu et al. (2019) employed the hedonic regression framework to collect determinants of rental prices and applied 6 machine learning techniques to forecast housing rent.

This research is the future work of Hu et al. (2019), aiming to adapt key determinants to suit the background of the Irish rental market and discover more effective techniques. A hedonic regression framework was employed to conduct the analysis. This research aims to investigate several key aspects of the Dublin rental market. Specifically, it aims to examine the current state of the rental market and assess whether renting creates a financial burden for Dublin residents. In addition, the study also aims to identify the main factors that influence rental prices in the city. Furthermore, it explores the effectiveness of various models in predicting rental prices and assesses whether deep learning models outperform traditional machine learning models in this case.

One of the main challenges in this project was the limited accessibility of rental datasets. To address this issue, listing data were scraped from Ireland's largest property website, Daft.ie, and additional feature data were collected from various reliable sources. Another significant challenge was the determination of an appropriate deep learning model architecture. To solve this, 3 deep learning housing price prediction models were selected from recent literature. The best-performing architecture was then identified after being evaluated using various metrics, and further hyperparameter tuning was employed to optimise the model's performance.

The major contributions of this project are threefold. First, it created a new dataset that captures important factors influencing rental prices in Dublin. Second, it applied the hedonic regression framework tailored to the Dublin rental market by selecting relevant and practical variables. Finally, the project fills a gap in the current literature by providing a comprehensive analysis of the determinants of rental prices in Dublin and predicting rental prices using machine learning and deep learning techniques.

However, this project was subjected to certain limitations. The relatively small dataset scraped (5000+ rows) limits its potential to capture complex patterns. Additionally, the short duration of data collection constrained the analysis of long-term patterns and seasonal variations in rental prices. Moreover, the use of simple deep learning models, combined with the application of random search for hyperparameter tuning, resulted in underperformance of deep learning models compared to machine learning techniques.

This paper was organised as follows. Section 2 examined studies that analysed the housing market based on key influencing factors and explored various methods used for predicting housing and rental prices. Section 3 outlined the methodology of this study, covering the process from data collection to data preprocessing, data analysis, and the application of prediction techniques. Section 4 presented the study's framework, introducing the three-layer architecture. Section 5 detailed the final implementation stage. Section 6 provided an in-depth analysis of the results and findings, supported by visualizations, statistical summaries, and charts. Lastly, Section 7 concluded the study and suggested potential directions for future research.

2 Related Work

Boeing and Waddell (2017) analysed 11 million Craigslist rental data in the US using statistics and visualisations. The proportion of income and rent per square foot with a threshold 30% was used to reflect the affordability. In addition, the geography was visualised to indicate the distribution of rents in the US. The results found that the high-rent properties are mainly concentrated on the west and east coasts, while the low-rent properties are concentrated in small towns and rural areas.

O'Hanlon (2011); Maguire et al. (2016) introduced the 12-month rolling hedonic regression model to identify the index to measure the housing price trend in Ireland. The factors, such as property types, locations, areas, and neighbourhood qualities, were considered in the model.

Lawless et al. (2018) proposed a new model to evaluate the rental indicators for 137 electoral areas across Ireland from 2007 to 2016. The results show a regional imbalance in rental returns in different environments. The zones closer to economic centres are more resilient in rental prices in various environments.

Shimizu et al. (2010); Goh et al. (2012) considered the factors such as areas, number of bedrooms, and locations in predicting rental prices.

Oduwole and Eze (2013) used the hedonic regression model to analyse the factors that influence the rental price in Abuja. The results reveal that the size of the room, the number of rooms and bathrooms, the accessibility to main roads, and the location of governments are the main factors contributing to the rental prices.

Neloy et al. (2019) used the size of an apartment, the number of bedrooms, the pet policy, the lobby condition, and the number of hospitals nearby to forecast the rent prices. The result shows the excellent performance of ensemble gradient boosting with an accuracy of 88.75%.

Wang et al. (2024) introduced a framework to analyse the relationships between the housing price and the school locations. Their research demonstrated the importance of school locations and qualities on housing prices in China.

Li et al. (2019) results evaluated that rental prices in different regions in Shanghai are affected by elements such as career, salary, size of the migrant population, and neighbourhood facilities and amenities.

Gan et al. (2016) explored the key determinants that affect the choice of rental property selection of migrant workers. The results revealed the importance of public facilities, the internal and external design of the building, prices, living quality and safety, and facilities management.

Kim et al. (2020) applied the LSTM model to predict housing prices in Busan, South Korea, utilizing 547,740 apartment transaction records while incorporating apartment characteristics and proximity to public rental housing. The results showed that proximity to Property Rental Housing (PRH)significantly impacts nearby property values at both city and neighbourhood levels.

Ming et al. (2020) analysed the key characteristics such as the housing area, orientation, way of paying rent, nearby transportation and structure. The results showed the XGBoost model achieved a better performance than Random Forest and LightGBM, in predicting the rent in Chengdu, China.

Thamarai and Malarvizhi (2020) utilised the dataset with factors of the number of bedrooms, transportation facilities, age of the property, schools, and shopping facilities to analyse and predict the housing price in Andhra Pradesh, India. The results revealed the outperformance of decision tree regression in predicting house prices.

Raju et al. (2023) applied the model random forest to forecast rental prices, considering factors house age, renovation conditions, size, number of rooms, and unit per square foot. The great accuracy of Random Forest in protecting rental prices was demonstrated.

Louzada et al. (2025) identified key rental price factors in São Carlos-SP, Brazil, and applied machine learning models including linear regression, SVR, Random Forest, XGBoost, and KNN. XGBoost showed the best performance based on the R2 metric. Additionally, the study also introduced a user-friendly website for rental estimation. However, the study was limited by its localised dataset, and future work aims to expand data diversity.

Xie et al. (2019) used various datasets for training, and compared the performance of machine learning techniques in forecasting the monthly rental prices. The results indicated that XGBoost and LightGBM models outperformed GBDT models.

Sakri et al. (2024) sed deep learning techniques and a tabular dataset with 2,930 entries and 82 variables to predict rental prices in Ames, Iowa. Applied deep learning techniques including DNN, CNN, LSTM, CNN+LSTM, CNN+LSTM+FCL, and CNN+LSTM+FCR. The structure of DNN was 3 hidden layers with the number of neurons equal to 40, 20, and 10, respectively. The solver Adams and the activation function ReLu were used in the model. The literation of the model was 40 times. As a result, the CNN+LSTM+FCL models outperformed others with the best R2 of 0.9615 and the smallest errors.

Chen (2024) compared the performance of the Linear Regression Model, Support Vector Regression and DNN in predicting the housing price in California. The data used in the study was tabular with 20,640 observations and 9 variables. The structure of RNN consisted of 4 hidden-layers with 128, 64, 32, and 16 neurons. The loss function used was MSE. The optimizer was Adam. The batch size was 64. The training epochs were set to 300, with early stopping triggered after 5 consecutive epochs without loss function improvement. The result demonstrated that DNN and SVR with RBF kernel outperformed the Linear Regression model.

Zhang et al. (2024) explored the performance of word-embedded techniques such as TF-IDF, Word2Vec, and BERT paired with proven algorithms including SVM, RF, GB, and DNN in predicting the house price. In addition to the tabular data, the house textual description data were included in the dataset. The structure of the DNN model contains 3 hidden layers with 100, 75, and 50 neurons. The hyperparameter explored included batch sizes ranging from 10 to 20, epoch values of 50, 75, and 100, learning rates of 0.0001, 0.001, and 0.01, activation functions between ReLU and sigmoid, and dropout rates ranging from 10% to 20%. The results demonstrated the best performance of the integrated model of DNNs and Word2Vec.

Truong et al. (2020) analysed the performance of 5 housing price prediction models, including lightGBM, random forest, hybrid regression, XGBoost, and stacked generalisation regression. The results showed the greater performance of hybrid regression and stacked generalisation regression in housing price prediction, with the ability to handle time complexities.

Ho et al. (2021) used 40,000 transaction data to compare the performance of support vector machine, gradient boosting machine, and random forest in housing price prediction in Hong Kong. The evaluation results of MSE, $RMSE$, and $MAPE$ showed the better performance of gradient boosting machine and random forest.

Hu et al. (2022) predicted the housing price in Shanghai and Wuhan in China, using the random forest model. The results not only demonstrated the great performance of random forest in housing price prediction, but also found the importance of the internal factors in housing price prediction.

Soltani et al. (2022) compared the performance of linear regression, random forest, decision tree, and gradient-boosted trees in housing price prediction. The results revealed the great performance of tree-based models. The best performing model among 4 techniques is random forest and gradient-boosted trees.

Mao (2024) applied random forest and gradient boosting to forecast the housing price. The results revealed the excellent performance of the model in predicting the fluctuations of the future prices of housing.

Maheshwari et al. (2024) highlighted the importance of factors such as locations, size, and number of bathrooms in predicting rental prices in Indian cities. The best performing models among all tested machine learning models were random forest and knn. The evaluation metrics used in the research include MSE,$RMSE$,R^2, and MAE.

Tandon (2024) utilised the listing data from 2019 to predict the housing prices in KUala Lumpur. The tested models including random forest, lightGBM, SVM, and GBR, showed great performance in housing price prediction. However, the best performing model among all tested models was GBR, with the lowest error and an excellent accuracy of 90%.

Wang et al. (2021) applied the Google Maps API to gather the property transaction data, facility locations, and satellite images. They proposed a new model to predict the housing price and compared its performance with traditional models, including XGBoost and LightGBM. The results demonstrated the better performance of the proposed model with lower prediction errors.

Chen et al. (2021) used the dataset from kaggle to compare the performance of 5 machine learning and deep learning models, including backpropagation neural network, linear regression, bayesian, deep neural network, and support vector machine. The result revealed the better performance of support vector machine, backpropagation neural network, and bayesian.

Embaye et al. (2021) predicted the housing price in Uganda, Tanzania, and Malawi using the survey data to evaluate the performance between the traditional OLS model and machine learning models. The result showed the better accuracy of machine learning models in housing price prediction.

Tan et al. (2024) proposed a housing price prediction model using a fully connected neural network, BERT, and MobileNet by employing the data with 50 features, description text and images. The results showed the great performance of the proposed model with an optimised MAPE of 5.57%, and demonstrated the effectiveness of multi-layer data and multiple neural networks in improving the accuracy of hosting price prediction.

Buranasiri and Laokulrach (2025) used a dataset of 983 rental listings in Pattaya, Thailand, incorporating variables such as distance to beach, property size, house age, room type, number of bedrooms and bathrooms, floor level, and sea view. The artificial neural network was developed and compared with stepwise multiple regression analysis. Results revealed that ANN outperformed the regression model with a higher forecasting accuracy.

3 Methodology

3.1 Datasets and Pre-processing

Determinants of the property rental price include three categories under the hedonic model, which are neighbourhood, location, and internal factors of the property Hu et al. (2019).

For the neighbourhood category, factors such as schoolsHu et al. (2019); Wang et al. (2024), healthcare facilities Neloy et al. (2019), entertainment facilities Gan et al. (2016); Thamarai and Malarvizhi (2020), and enterprise park Li et al. (2019) locations were selected. The school types considered include primary school and post-primary education, downloading from gov.ie [4]. The hospital data was identified on the HSE website [5]. The shopping centres and enterprise parks datasets were collected using the Google Maps API.

The pre-processing of the neighbourhood data involved filtering out facilities that were not located in Dublin.

Regarding location category, public transportation is the primary factor influencing property rental prices Hu et al. (2019); Li et al. (2019); Ming et al. (2020); Thamarai and Malarvizhi (2020). In Dublin, Luas and trains are the two major public transport options, due to their punctuality and stability. The detailed transportation data was gathered using the Google Maps API, including station names, addresses, and coordinates.

The first step in preprocessing transportation data was to merge two datasets to create a new transport stops dataset. Next, duplicate stops serving both Luas and train services were removed.

Previous studies also demonstrated the importance of internal factors such as property types and sizes Boeing and Waddell (2017); Lima Holanda (2018); Hu et al. (2019), locations Hurley and Sweeney (2024); Goh et al. (2012), number of bedrooms Neloy et al. (2019); Lima Holanda (2018), and number of bathrooms Lima Holanda (2018); Oduwole and Eze (2013). Therefore, 8 internal factors were scraped from Ireland's largest listing website daft.ie using the Python library 'daftlistings' [6]. The collected variables included the address, coordinates, price, the BER ratings, the property type, the number of bedrooms and bathrooms, and the published time. In addition, rental data were collected monthly from November 2024 to April 2025 to expand the timespan of data.

There were 6 steps in pre-processing the main dataset. First, duplicate listings reposted by landlords or agents were removed. The weekly rental prices were converted to the monthly rental prices.

Second, Dublin was divided into 12 constituency boundaries, following a well-defined and consistent geographical framework that aligns with the population distribution. The Geojson data for the electoral zone boundaries was obtained from the government's open

[4] https://www.gov.ie/en/collection/63363b-data-on-individual-schools/
[5] https://www.geohive.ie/datasets/feb34881088341bbbf80d86af6a4f333_0/about
[6] https://github.com/AnthonyBloomer/daftlistings

data portal [7]. Dummy Variables were created to represent each region in order to further the analysis of geographic influences on rental prices.

Third, the BER ratings and property types were converted to numerical levels. There are 15 levels of BER, ranging from A1 to G, where A1, the highest energy rating, is assigned a value of 1, and G, the lowest rating, is assigned a value of 15. The medium value was given to the properties missing the BER value. Similarly, the property types, House, Apartment and Studio Apartment, were assigned values 1, 2, and 3, respectively, based on their financial value.

Fourth, missing values for the number of bathrooms and bedrooms were imputed with the most appropriate or relevant values, determined through online research of similar properties.

Fifth, the closest facilities were found for each property. The distances (in km) between each property and the nearest facilities were calculated.

To complete the processing of the main dataset, the cleaned data was stored in the PostgreSQL database for efficient querying.

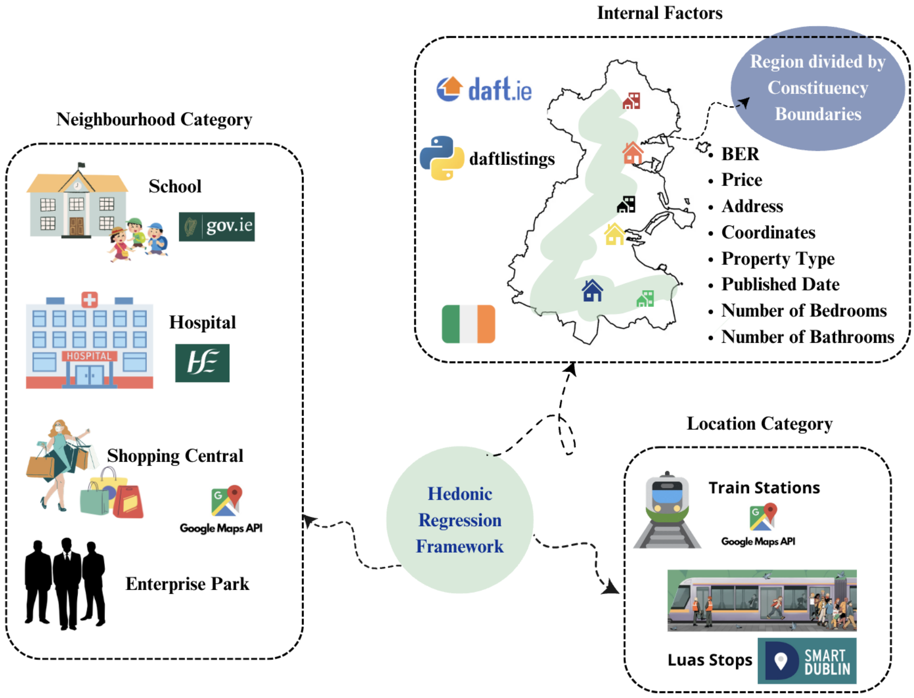

Figure 1: Hedonic Regression Framework

3.2 Data Analysis

The data analysis section was designed to address the first three research objectives, guided by the hedonic model framework. A combination of statistical analysis and data visualisation techniques was employed to address the first and third objectives.

First, distribution and box plots of the rental prices were plotted to visualise the overall distribution and identify outliers.

[7]https://data-osi.opendata.arcgis.com/datasets/a37ad6a3a6ff47e4a5a0ff313b418448_0/explore

Second, the distribution plots of distance to facilities were generated to evaluate the accessibility to nearby facilities.

Third, the heat maps of mean and median rental prices across various regions in Dublin were generated to explore regional differences in rental prices.

Fourth, the 15-minute walking distance concept, introduced by Hu et al. (2019), was applied to the research. The average walking speed used by Google Maps, 5 km/h, was used to measure the distance. As a result, the 15-minute walking distance was equivalent to 1.25 km. Therefore, the properties were grouped into two categories, within and beyond 1.25 km. Furthermore, the Shapiro-Wilk test, Mann-Whitney U test, and T-test were utilised to measure the statistical significance of rental prices between two groups, with a significance level of 0.05.

Fifth, box plots were employed to visualise the distribution of property types, BER ratings, and the number of bedrooms and bathrooms across different levels.

Lastly, the plot between rental prices and times was generated to evaluate the time trends and potential fluctuations in the rental market.

The second objective was addressed by applying the Rent-to-Income Ratio Boeing and Waddell (2017). The ratio was obtained by dividing the median rent by the median income across various populations. The threshold, 30%, was used to determine whether the rental burden exists, with results above the threshold being true.

3.3 Machine Learning and Deep Learning Algorithms

The predictive techniques applied in the research were selected from the review of papers. Random Forest Regression Thamarai and Malarvizhi (2020); Raju et al. (2023), XGBoost Ming et al. (2020); Louzada et al. (2025), SVR Hu et al. (2019); Chen (2024), LightGBM Xie et al. (2019), and GBR Hu et al. (2019); Neloy et al. (2019), were applied in the study.

The deep learning technique selected for the research was Deep Neural Networks (DNNs), due to its great performance on the Tableau data. Three DNN architectures were selected to find the best configurations. The first model was based on the study of Sakri et al. (2024). The second model was based on the research of Chen (2024). The last architecture was based on the study of Zhang et al. (2024).

The first architecture contained 3 hidden layers with 40, 20, and 10 neurons. The optimiser, activation function used in the first architecture was Adam and ReLU. The number of epochs was 40.

The second model, however, had 4 hidden-layers with 128, 64, 32, and 16 neurons. The loss function and optimiser used were MSE and Adam. The batch size was 64. The epochs were set to 300, with early stopping triggered after 5 continuous epochs without MSE improvement.

The last model used a grid search to find the best parameters. The number of layers for the third model was 3, with 100, 75, and 50 neurons. The evaluated bath sizes were 10 and 20. The number of epochs was selected between 50, 75, and 100. The learning rate was chosen from 0.01, 0.001, and 0.0001. The two activation functions compared were ReLU and sigmoid. The evaluated dropout rates ranged from 10% and 20%.

All three models were trained on the pre-processed data, and the best-performing model was then further tuned using a random walk strategy to optimise its parameters for predicting property rental prices. The Python packages sklearn and tensorflow were used in implementing deep learning techniques.

To compare the performances of predictive models in forecasting property rental price, the metrics R^2, MAE, $\%MAE$, MSE, $RMSE$, and $\%RMSE$. R^2 was used in the project to measure how good the models fit the data. MAE and $RMSE$, as well as their percentages, were used to assess the precision of the models by measuring the gaps between the actual and predicted values. $RMSE$ and its percentage, however, are more sensitive to outliers and large errors.

In addition, histograms of feature importance generated through the machine learning modelling were used to identify the hidden variables that contribute to Dublin rental prices.

4 Design Specification

The summarised architecture of this project contains 3 layers. The first layer was data collection and preprocessing under the hedonic regression framework. The main dataset contained the internal factors of rental properties, the neighbourhood data, and the public transport data were collected. The data processing process contains the region identification, BER and property type encoding, handling missing values, and nearest facility distance calculation.

The second layer was data analysis and visualisation. The statistics and visualisations, such as the bar chart, heat map, and box plot, were found to perform a comprehensive exploratory data analysis of the rental price in Dublin. The first three research objectives were addressed at this layer.

The last layer involved predicting rental prices using deep learning and machine learning techniques. This process included exploring the data distribution and applying a log transformation to reduce skewness in the variables. Hyperparameter tuning was performed on the machine learning models to optimise their performance and identify the most effective configurations. Additionally, the predictive performance of the initial 3 DNN architectures was evaluated. A random Walk strategy was then applied to the best-performing architectures to further fine-tune parameters and enhance model performance. The last two research objectives were addressed at this layer.

To evaluate the performance of each model, the metrics R^2, MAE, $\%MAE$, $RMSE$, $RMSE$, and $\%RMSE$ were utilized.

The hardware used in this project consisted of a MacBook Pro with an M1 chip and 16 GB of RAM. The software environment included Python 3, Jupyter Notebook, Google Maps API, and PostgreSQL. The implementation leveraged various libraries, including Pandas, NumPy, Scikit-learn, TensorFlow, daftlistings, Matplotlib, and Seaborn.

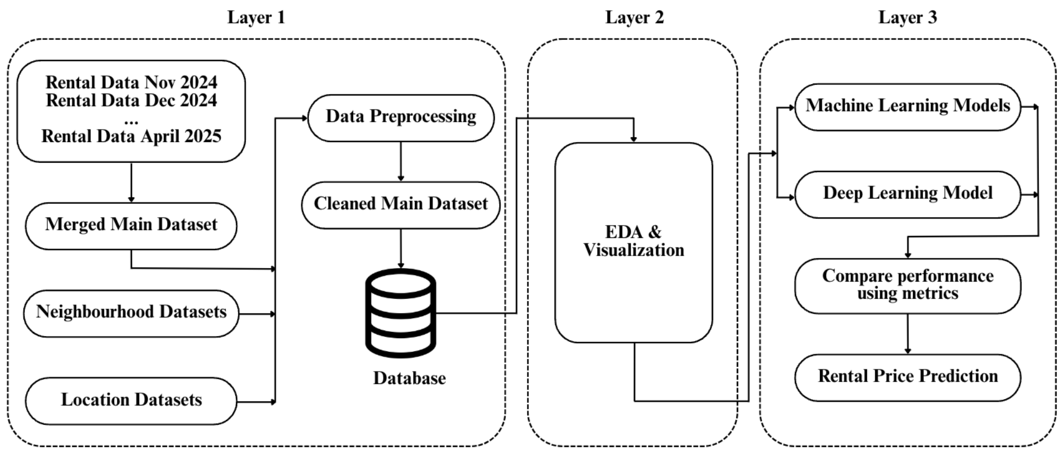

Figure 2: 3-layer Architecture

5 Implementation

5.1 Data Analysis

Firstly, the distribution and box plot of the price and distance to facilities are plotted in Figure 3 (a). The right skew of all distributions was observed, and a large number of outliers in the dataset beyond the upper bound of the box plot were found. The statistics of the rental prices across all properties and regions were shown in Figure 4.

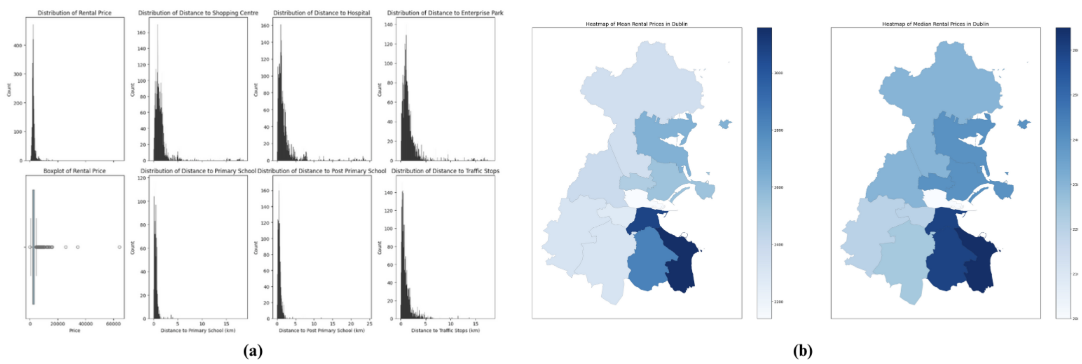

Figure 3: (a) Distributions and Box Plot of Rental Prices and Distance to Facilities (b) Regional Price Differences

	Mean	Std.	Lower Bound	Q1	Median	Q3	Upper Bound
All	2660.4	1675.6	450.0	1950.0	2400	2950.0	4450.0
Dún Laogharie	3157.4	3455.1	790.0	2290.0	2650.0	3250.0	4750.0
Dublin Bay South	3076.3	1875.0	500.0	2000.0	2600.0	3500.0	5000.0
Dublin Rathdown	2829.3	932.4	700.0	2200.0	2600.0	3245.0	4745.0
Dublin Final East	2629.1	1226.1	530.0	2030.0	2400.0	2759.0	4259.0
Dublin Bay North	2549.6	1158.1	548.5	2048.5	2400.0	2900.0	4400.0
Dublin North West	2493.2	859.3	500.0	2000.0	2400.0	2700.0	4200.0
Dubin West	2389.2	870.0	475.5	1975.5	2300.0	2652.5	4152.5
Dublin Fingal West	2356.2	812.4	450.0	1950.0	2300.0	2550.0	4050.0
Dublin Mid West	2322.2	915.7	372.3	1872.3	2200.0	2651.5	4151.5
Dublin South West	2319.9	750.1	350.0	1850.0	2244.0	2737.5	4237.5
Dublin South Central	2278.8	660.6	350.o0	1850.0	2200.0	2565.8	4065.8
Dublin Central	2140.8	699.1	200.0	1700.0	2000.0	2500.0	4000.0

Figure 4: Statistics of Rental Price

Second, the mean and median rental price variations across regions were analysed through a heat map (Figure 3 (b)).

Third, the 15-minute walking distance to each facility was examined. The Shapiro-Wilk test was applied to assess the normality of the data distribution. The results showed that the data did not follow a normal distribution. Therefore, the U test was used to assess statistical significance. The results are presented in Figure 5.

	p-value	Significant
Traffic Stops	0.175	False
Enterprise Park	3.546e-22	True
Shopping Central	0.824	False
Hospital	0.022	True
Primary School	0.243	False
Post Primary School	0.975	False

Figure 5: Rental Price Differences: Inside vs. Outside 15-Minute Walking Distance

Fourth, the internal factors were analysed. The box plots of the price of each internal factor were drawn in Figure 6 (a).

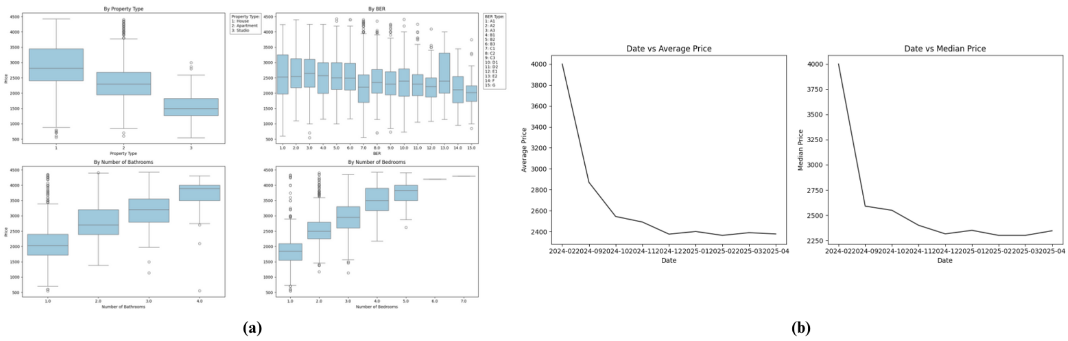

Figure 6: (a) Internal Factor Box plots (b) Time vs Rental Price

Fifth, the relationship between time and rental price was examined. The data were grouped by month, and the mean and median rental prices were calculated for each period. These values were plotted as a time series to illustrate changes in rental prices over time (Figure 6 (b)).

Last, according to data from Ireland's Central Statistics Office [8], the median nominal household disposable income in 2024 was €58,922, equivalent to approximately €4910 per month. A breakdown of household income by composition was provided in the Figure 7. The Rent-to-Income Ratio (RIR) was calculated for each household composition, where a ratio greater than 30% was considered to be rent-burdened.

	Median Annual	Median Monthly	RIR (%)
All	58,922	4,910.17	48.88
1 Adult Aged 65+	20,067	1672.25	-43.52
1 Adult Aged <65	32,000	2666.67	90
2 Adults, At Least 1 Aged 65+	44,737	3728.08	64.38
2 Adults, Both aged <65	64,403	5366.92	44.72
3 or More Adults	85,716	7143.00	33.60
1 Adult with Children Aged Under 18	38,929	3244.08	73.98
2 Adults with 1-3 Children Aged Under 18	70,249	5854.08	41.00
Other Households with Children Aged Under 18	84,590	7049.17	34.05

Figure 7: Rental Burden Analysis

5.2 Machine Learning and Deep Learning

As the data analysis revealed skewed distributions in the variable price and distance to facilities, the log transformations were applied to normalise the variables. In addition, outliers were removed to make sure the data was ready for modelling. To prepare for modelling, the dataset was split into training and testing parts using an 80/20 ratio, and the corresponding subsets were saved as CSV files for further use.

[8] https://www.cso.ie/en/releasesandpublications/ep/p-silc/surveyonincomeandlivingconditionssilc2024/householdincome/

In terms of machine learning techniques, hyperparameter tuning was performed to identify the best-performing parameters. To enhance efficiency, the random walk strategy was used to explore the parameters. As a result, the optimal parameters for the Random Forest included 500 estimators, a minimum sample leaf size of 1, a minimum sample split size of 2, and a maximum feature selection method set to 'sqrt'. For the XGBoost model, the best configuration consisted of 500 estimators, a learning rate of 0.05, a subsample ratio of 1, a maximum depth of 7, a lambda value of 5, an alpha value of 0.1, and a column subsample ratio of 0.6. For the LightGBM, the best parameter consisted of a 0.7 subsample, 50 number of leaves, 500 estimators, 5 minimum child samples, -1 maximum depth, 0.05 learning rate, and 0.8 columns subsample ratio. For GBR, the optimal parameters consisted of 0.8 of subsample, 500 estimators, 5 minimum sample split, 4 minimum samples leaf, 5 maximum depth, and 0.05 learning rate. Additionally, a feature importance histogram (Figure 8) was plotted to visualise the contribution of each independent features to the rental prices.

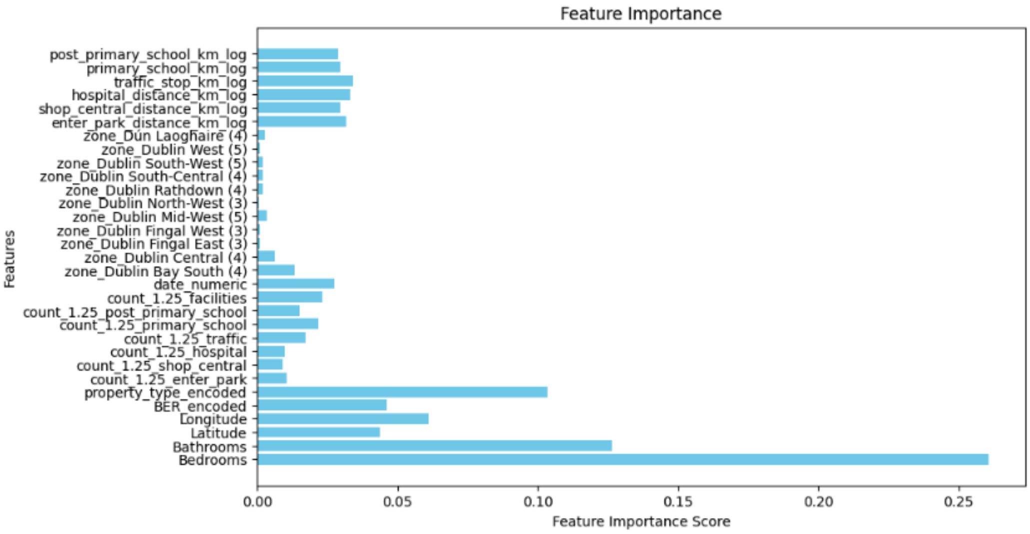

Figure 8: Feature Importance

For the deep learning technique, the best-performing architecture was based on the study of Zhang et al. (2024), which achieved the lowest prediction error and a model fitting rate exceeding 70%. Consequently, further fine-tuning was conducted on this architecture to optimise its prediction performance. The final best-performing DNN model consisted of 2 hidden layers with 100, 75, and 50 neurons respectively, ReLU activation functions, 2 dropout layers with a dropout rate of 0.2, and was trained for 100 epochs.

The metrics results of R^2, MAE, $\%MAE$, $RMSE$, and $\%RMSE$. R^2 was generated to show the performance of all models, and the results were shown in Figure 9.

	Random Forest	XGBoost	RNNs	SVR	LightGBM	GBR
R-Squared	0.77	0.78	0.74	0.21	0.79	0.79
MAE	240.10	233.82	266.15	554.26	234.97	235.80
MSE	116752.79	108867.02	132662.30	505483.76	106189.96	10675.21
RMSE	341.69	329.95	364.23	710.97	325.87	326.73
%RMSE	13.96	13.48	14.89	29.06	13.32	13.35
%MAE	10.61	10.45	12.16	24.33	34.61	34.5

Figure 9: Performance of Technologies

6 Evaluation

In terms of the answers to the first question, the rental market in Dublin is characterised by high and unevenly distributed rental prices. The data revealed a right-skewed distribution in rental prices across properties (Figure 3 (a)), with an average monthly rent of €2,600.4 and a median of €2,400 (Figure 4). Regional disparities were evident, with higher rental prices concentrated along Dublin Bay, particularly in southern regions such as Dún Laoghaire, Dublin Bay South, and Dublin Rathdown (Figure 3 (b) and Figure 4). A temporal analysis (Figure 6 (b)) also indicated a noticeable decline in rental prices between February and December 2024, followed by slight fluctuations.

For the second question, the analysis suggested a significant rental burden for Dublin residents. According to Figure 7, the Rent-to-Income Ratio exceeded the 30% threshold for all household types, which indicates rent stress. The burden is especially pronounced among households such as single adults above 65, older couples with at least one person aged above 65, and single-parent families with young children.

In terms of the answers to the third question, several external and internal factors significantly influence rental prices. Proximity to enterprise parks and hospitals showed statistically significant effects on rent levels, with higher prices for properties within a 15-minute walking distance (Figure 5). Internally, property type, BER, and the number of bedrooms and bathrooms were key features influencing the rental prices, as visualised in Figure 6 (a). These results highlight the effect of accessibility and housing characteristics on rent pricing.

6.1 Machine Learning and Deep Learning

In addition to data analysis through visualisations, the answer to the third question was extended by the Figure 8 generated from machine learning models. The results revealed hidden relationships between rental prices and various features, including Spatial variables (latitude and longitude), the number of facilities within a 15-minute walking distance, and the distance to the nearest facilities.

Figure 9 indicates that the machine learning models outperformed the deep learning model, achieving lower prediction errors and high model fit. Among all models, XGBoost, LightGBM, and GBR performed well, with LightGBM demonstrating the best performance. However, the deep learning model was not outperfromed the machine learning in this research.

6.2 Discussion

Summarising the previous two sections, the findings were as follows:

Firstly, Dublin was identified as facing a severe rental burden, as indicated by a high Rent-to-Income Ratio.

Secondly, rental prices vary significantly in different regions in Dublin, with the highest rental prices concentrated in the east coastal areas.

Thirdly, statistical significance was found for the properties within and beyond 15-minute walking distance to hospitals and enterprise parks. This reflects a stronger preference for proximity to employment opportunities and healthcare services, which is consistent with research by Li et al. (2019); Neloy et al. (2019).

Fourthly, the internal factors including property type, number of bathrooms, number of bathrooms, and BER rate, were shown to significantly influence rental prices. This finding was consistent with the studies by Oduwole and Eze (2013); Neloy et al. (2019); Ming et al. (2020); Maheshwari et al. (2024).

Fifthly, the hidden relationships between rental prices and spatial variables, including coordinates and proximity to multiple facilities, were revealed via feature importance histogram, which demonstrated the feature importance of neighbourhood and location variables, aligning with the study of Wang et al. (2024); Gan et al. (2016); Hu et al. (2019); Li et al. (2019); Ming et al. (2020).

Sixthly, time trends were observed for both mean and median rent prices, with rental prices fluctuating over time.

Lastly, the outcomes of machine learning models, particularly LightGBM, was found to be better than the selected deep learning model in predicting rental prices in Dublin in this research. In addition, the performance of LightGBM, GBR, and XGboost demonstrated the excellent performance of tree-based models in capturing non-linear relationships, and thus, along with the study of Hu et al. (2019); Duerr et al. (2018); Rendall et al. (2017).

However, the bad performance of the deep learning model compared with machine learning models was surprising. The small size of the dataset was a potential reason. Deep learning models typically require large columns of data to effectively capture complex patterns. In this study, the limited number of observations might have led to overfitting, even with dropout layers implemented to mitigate it. Therefore, the larger dataset with a larger time span and unstructured data types such as images and texts should be collected in the future to improve the performance. Additionally, the random search process used for hyperparameter tuning may not have been comprehensive enough to identify the optimal configuration for the deep learning model. Therefore, more advanced or targeted search methods, such as Bayesian optimisation, could be used to improve the performance. Moreover, the deep learning model selection was relatively limited, relying mainly on the basic deep neural network. This choice may not have been flexible enough to capture the nuanced relationships between features in the dataset. Therefore, more advanced deep learning techniques such as the embedded models CNN+LSTM+FCR and CNN+LSTM+FCL Sakri et al. (2024) could be explored in future work to improve performance and robustness.

7 Conclusion and Future Work

This study proposed a model that integrated the hedonic regression model with predictive techniques to analyse Dublin rental prices, aiming to identify key characteristics and the most effective forecasting model. The evaluation metrics used in this study include R^2, MAE, $\%MAE$, MSE,$RMSE$, and $\%RMSE$.

The findings of this study effectively address the five research objectives. Data visualisations and rental burden analysis were conducted to examine Dublin's rental market and assess whether renting is a burden on residents. The finding revealed significant regional disparities in rental prices, with the highest prices observed in the southern areas and coastal zones of Dublin. Moreover, a considerable rental burden was identified, particularly affecting elderly residents and single-parent households.

To identify the key factors influencing rental prices, the study employed visual analytics and machine learning modelling. The results revealed that the features, including property types, number of bedrooms and bathrooms, BER rating, and proximity to various facilities, highly influenced the rental prices. Additionally, spatial features and the number of nearby facilities within a 15-minute walking distance were found to contribute to rental price predictions.

Finally, to determine the most effective rental price prediction model and assess the comparative performance of machine learning and deep learning techniques, several models were evaluated using selected performance metrics. Machine learning models, particularly LightGBM, outperformed deep learning models (DNNs), achieving better data fitting and the lowest prediction errors, with an R^2 of 0.79, MSE of 106189.96, MAE of 234.97, $RMSE$ of 325.87, $\%RMSE$ of 13.32%, and $\%MAE$ of 34.61%.

Several limitations were identified in this research, including the relatively small dataset, the use of basic deep learning models, and the application of random search for hyperparameter tuning. To address their limitations, future work is proposed, including expanding the dataset via web-scraping to cover a broader time range and integrating additional data types such as images and textual descriptions. In addition, advanced deep learning techniques such as the embedded models CNN+LSTM+FCR and CNN+LSTM+FCL should be explored to predict rental prices. Furthermore, more advanced hyperparameter tuning algorithms, including Bayesian optimisation, should be considered to enhance the model's performance.

8 Appendix

Term/Acronym	Definition/Description
Hedonic Regression	A method to estimate how different factors influence housing prices.
Web Scraping	Automated extraction of data from websites.
Spatial Analysis	Analysis of geographic or location-based data.
Proximity Measures	Distances from properties to nearby amenities.
Google Maps API	A tool to access map data, locations, and distances programmatically.
Central Statistics Office (CSO)	Ireland's official provider of economic and population statistics.
Rent-to-Income Ratio (RTR)	Share of income spent on rent; measures affordability.
Building Energy Ratings (BER)	Energy efficiency score for buildings.
Random Forest (RF)	A machine learning model using multiple decision trees.
Support Vector Regression (SVR)	A regression method based on Support Vector Machines.
XGBoost	An optimised gradient boosting algorithm.
LightGBM	A fast, efficient gradient boosting framework.
Gradient Boosting Regressor (GBR)	A model combining decision trees using boosting.
Recurrent Neural Networks (RNNs)	A deep learning model for sequential data.
R^2	Indicates how well features explain variation in target values.
MSE	Mean of squared prediction errors.
RMSE	Square root of MSE; error in original units.
%RMSE	RMSE as a percentage of actual values.
%MAE	MAE as a percentage of actual values.

Figure 10: Table of Definitions, Terms, and Acronyms

References

Boeing, G. and Waddell, P. (2017). New insights into rental housing markets across the united states: Web scraping and analyzing craigslist rental listings, *Journal of Planning Education and Research* **37**(4): 457–476.

Buranasiri, J. and Laokulrach, M. (2025). Appropriate rental price prediction for condominiums in pattaya, thailand using artificial neural network approach.

Chen, A. (2024). Deep learning in real estate prediction: An empirical study on california house prices, *The National High School Journal of Science* .

Chen, Y., Xue, R. and Zhang, Y. (2021). House price prediction based on machine learning and deep learning methods, *2021 International conference on electronic information engineering and computer science (EIECS)*, IEEE, pp. 699–702.

Duerr, I., Merrill, H. R., Wang, C., Bai, R., Boyer, M., Dukes, M. D. and Bliznyuk, N. (2018). Forecasting urban household water demand with statistical and machine learning methods using large space-time data: A comparative study, *Environmental Modelling & Software* **102**: 29–38.

Embaye, W. T., Zereyesus, Y. A. and Chen, B. (2021). Predicting the rental value of houses in household surveys in tanzania, uganda and malawi: Evaluations of hedonic pricing and machine learning approaches, *Plos one* **16**(2): e0244953.

Gan, X., Zuo, J., Chang, R., Li, D. and Zillante, G. (2016). Exploring the determinants of migrant workers' housing tenure choice towards public rental housing: A case study in chongqing, china, *Habitat International* **58**: 118–126.

Goh, Y. M., Costello, G. and Schwann, G. (2012). Accuracy and robustness of house price index methods, *Housing Studies* **27**(5): 643–666.

Ho, W. K., Tang, B.-S. and Wong, S. W. (2021). Predicting property prices with machine learning algorithms, *Journal of Property Research* **38**(1): 48–70.

Hu, C., Huang, R. and Li, H. (2022). Prediction and analysis of rental price using random forest machine learning technique take shanghai and wuhan for example, *2022 International Conference on mathematical statistics and economic analysis (MSEA 2022)*, Atlantis Press, pp. 587–593.

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M. and Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies, *Land use policy* **82**: 657–673.

Hurley, A. K. and Sweeney, J. (2024). Irish property price estimation using a flexible geo-spatial smoothing approach: What is the impact of an address?, *The Journal of Real Estate Finance and Economics* **68**(3): 355–393.

Kim, H., Kwon, Y. and Choi, Y. (2020). 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* **12**(18): 7520.

Lawless, M., McQuinn, K. and Walsh, J. (2018). Identifying rent pressures in your neighbourhood: A new model of irish regional rent indicators, *The Economic and Social Review* **49**(1, Spring): 73–92.

Li, H., Wei, Y. D. and Wu, Y. (2019). Analyzing the private rental housing market in shanghai with open data, *Land Use Policy* **85**: 271–284.

Lima Holanda, V. (2018). Delivering social housing: An overview of the housing crisis in dublin.

Louzada, F., de Lacerda, K. J. C. C., Ferreira, P. H. and Gomes, N. D. (2025). Smart renting: Harnessing urban data with statistical and machine learning methods for predicting property rental prices from a tenant's perspective, *Stats* **8**(1): 12.

Maguire, P., Miller, R., Moser, P. and Maguire, R. (2016). A robust house price index using sparse and frugal data, *Journal of Property Research* **33**(4): 293–308.

Maheshwari, A., Ranka, M., Malhotra, A., Mishra, R. K. and Basha, M. S. A. (2024). Assessing machine learning models for predictive analytics in urban rental markets: Rent price forecasting in india, *2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, IEEE, pp. 1–9.

Mao, M. (2024). A comparative study of random forest regression for predicting house prices using, *2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Atlantis Press, pp. 619–626.

Ming, Y., Zhang, J., Qi, J., Liao, T., Wang, M. and Zhang, L. (2020). Prediction and analysis of chengdu housing rent based on xgboost algorithm, *Proceedings of the 3rd International Conference on Big Data Technologies*, pp. 1–5.

Neloy, A. A., Haque, H. S. and Ul Islam, M. M. (2019). Ensemble learning based rental apartment price prediction model by categorical features factoring, *Proceedings of the 2019 11th International conference on machine learning and computing*, pp. 350–356.

Oduwole, H. and Eze, H. (2013). A hedonic pricing model on factors that influence residential apartment rent in abuja satellite towns, *Mathematical Theory and Modeling* **3**(12): 65–74.

O'Hanlon, N. (2011). Constructing a national house price index for ireland., *Journal of the Statistical & Social Inquiry Society of Ireland* **40**.

Raju, R., Neyaz, A., Ahmed, A. and Singh, A. (2023). Machine learning for rental price prediction: Regression techniques and random forest model, *Available at SSRN 4587725* .

Rendall, R., Pereira, A. C. and Reis, M. S. (2017). Advanced predictive methods for wine age prediction: Part i–a comparison study of single-block regression approaches based on variable selection, penalized regression, latent variables and tree-based ensemble methods, *Talanta* **171**: 341–350.

Sakri, S., Ali, Z. and Ismail, N. H. A. (2024). Assessment of the hybrid deep learning models and hedonic pricing model for house price prediction, *2024 Seventh International Women in Data Science Conference at Prince Sultan University (WiDS PSU)*, IEEE, pp. 127–133.

Shimizu, C., Nishimura, K. G. and Watanabe, T. (2010). Housing prices in tokyo: A comparison of hedonic and repeat sales measures, *Jahrbücher für Nationalökonomie und Statistik* **230**(6): 792–813.

Soltani, A., Heydari, M., Aghaei, F. and Pettit, C. J. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms, *Cities* **131**: 103941.

Tan, H., Su, T., Wu, X., Cheng, P. and Zheng, T. (2024). A sustainable rental price prediction model based on multimodal input and deep learning—evidence from airbnb, *Sustainability* **16**(15): 6384.

Tandon, R. (2024). The machine learning based regression models analysis for house price prediction, *International Journal of Research and Analytical Reviews (IJRAR)* **11**(3): 296–305.

Thamarai, M. and Malarvizhi, S. (2020). House price prediction modeling using machine learning., *International Journal of Information Engineering & Electronic Business* **12**(2).

Truong, Q., Nguyen, M., Dang, H. and Mei, B. (2020). Housing price prediction via improved machine learning techniques, *Procedia Computer Science* **174**: 433–442.

Wang, P.-Y., Chen, C.-T., Su, J.-W., Wang, T.-Y. and Huang, S.-H. (2021). Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism, *IEEE Access* **9**: 55244–55259.

Wang, S., Li, Y. and Gao, Z. (2024). School-housing nexus in urban china: a comparative study of the effect of school districts on commercial and danwei housing prices, *Applied Spatial Analysis and Policy* **17**(2): 417–438.

Xie, Y., Xiang, W., Ji, M., Peng, J. and Huang, Y. (2019). An application and analysis of forecast housing rental based on xgboost and lightgbm algorithms, *Computer applications and software* **36**(9): 151–155.

Zhang, H., Li, Y. and Branco, P. (2024). Describe the house and i will tell you the price: House price prediction with textual description data, *Natural Language Engineering* **30**(4): 661–695.