National College of Ireland

Predicting Ireland House Prices with Deep Learning Techniques - A Comparative Study

MSc Research Project
MSc Data Analytics

Nithish Christopher
Student ID: x23116099

School of Computing
National College of Ireland

Supervisor: Anu Sahni

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Nithish Christopher
Student ID:	x23116099
Programme:	MSc Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Anu Sahni
Submission Due Date:	12/12/2024
Project Title:	Predicting Ireland House Prices with Deep Learning Techniques - A Comparative Study
Word Count:	5351
Page Count:	19

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:	Nithish Christopher
Date:	29th January 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):	

Predicting Ireland House Prices with Deep Learning Techniques - A Comparative Study

Nithish Christopher
x23116099

Abstract

The Republic of Ireland is one of the developed countries in the world. The rapid growth of the housing market is a significant factor in Ireland's economic development. In recent years, the housing crisis and the increasing house prices have become prominent issues in Ireland, as reported by the Irish Central Newsletter. This research aims at developing a predictive model for the gradual increase in house prices in the Republic of Ireland using deep learning, focusing on historical data analysis from recent years. The data for this study was collected from the Property Price Register Ireland dataset, covering the period from 2010 to 2021. The research focuses on developing and comparing deep learning models, which includes Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), Long Short-Term Memory Networks (LSTM) and Recurrent Neural Networks (RNN), to identify the best model for predicting house prices. Among these models, the Artificial Neural Networks (ANN) performed the best, based on the RMSE value.

1 Introduction

The republic of Ireland, facing biggest challenge in the form of a housing crisis and rising house prices over the past few years. One of the reasons for this is the increasing population because of development in technology and healthcare with more features and upgrades. According to the Social Justice Ireland census report, house prices have been rising not only in Dublin but also in other counties. Based on the mentioned details affecting various groups, including new immigrants and students, who are struggling with high rental costs. Additionally, citizens of Ireland are also experiencing disturbances in their peaceful life due to the housing crisis and increase in price issue. These mentioned factors were the reasons that motivated this research.

The research focuses on complete Ireland house price prediction. This research includes fluctuation of prices in past years, seasonal trends of prices, category of price ranges. This analysis also helps in better understanding of Ireland housing market and provides knowledge about price fluctuation. This research will be effective for Ireland house price prediction. This research is based on Property Price Register Ireland dataset on Kaggle, containing 2010 to 2021 data. Advanced techniques are implemented in this research for better feasible prediction.

1.1 Research question

How can different deep learning models, including ANN, RNN, CNN, LSTM be optimized for accurate and scalable house price predictions in Republic of Ireland?

- **Sub- RQ:** How does the predicted house price value from optimized neural network models compare to the actual values in Ireland's housing market?

A recent work which utilized the same concept, implemented machine learning and Deep learning techniques, particularly ANN and Auto ML techniques, this research was conducted by Aditiya Ravindra Ved ed.al (2024). Here, they mentioned to do future work on improving models by exploring advanced deep learning techniques with detailed data preprocessing, addressing time-based trends. The above-mentioned suggestions in the existing works were taken into account in this study.

1.2 Document Structure

This research was performed in a well structure format and documented in detail. The beginning of the abstract gives complete overview of the research objective, introduction mentioning detailed overview of the purpose to do this research for republic of Ireland, mentioned strongly on base study of this research and detailed related work, research methodology of the study, Implementation process, in the implementation, data cleaning, data preprocessing and feature engineering, model implementation and evaluation metrics like RMSE, MAE, MSE were validated, provided the predicted value for suitable model and concluded the research.

2 Related Work

2.1 Advanced Data Preprocessing and Insightful Analysis for Optimizing House Price Predictions

This research conducted by (Ved and Gupta; 2024) strongly discussed data preprocessing such as handling missing values, dealing with outliers in detail, encoding of categorical data on Ames city in the United Staes. The authors applied AutoML technologies and ANN, those preprocessing are effectively used on applying these models. It mainly compares traditional models like Ridge Regression with advanced techniques such as Artificial Neural Networks (ANN) and AutoML technologies. At the same time, the writers focused on limited sequential technique. The paper not only focused on handling sequential data, but it compares traditional and advanced models. It doesn't explain in detail why some models performed better than others in certain situations, especially for neural network (ANN). This research is more helpful for my research in Ireland house price prediction because it provides detailed essential data preprocessing techniques and EDA. These are much important for implementing large data with advanced techniques are ANN, RNN, CNN, LSTM. Additionally, the comparison of traditional machine learning models with advanced approaches like ANN gave me a clearer understanding on how different methods can be implemented, assisting me with a design on balanced and effective predictive value.

2.2 Discussing House Price Prediction on Advanced Deep Learning

In this research aim to develop accurate and effective models for predict house price by analyzing with valuable insights. In this analysis they used Multiple Linear Regression, SVR, ANN, XGBoost. Implementing of those techniques to they identified key factors of influencing house prices. Compared with those models to be identified Gradient Boosting (XGBoost) gave superior accuracy. Disadvantage of the analysis is limited dataset scope, minimal interpretability for ANN and XGBoost. some feature selection also identified these are challenges on model performance and research focused on single location. This complete research was done by (Saini et al.; 2023).

This study aims at comparing the performance of Na¨ıve Base, Random Forest and ANN to predict the house price prediction based various collected data. In this research they focused on detailed comparative performance analysis. In this research ANN got top accuracy, they had detailed preprocessing as well. In this research also they covered smaller datasets, lacking evaluation metrics like MAE and RMSE. This research was conducted by (V et al.; 2024).

In this study aims to predict real time house price using of XgBoost algorithm, Linear regression. In this analysis they finalized performance based on RMSE values to concluded with XG-boost is the high performance. But in this research if they proceed with smaller data and for one small region. The study done by (Joseph et al.; 2024) requires regression comparison with multiple regression model.

This study extends the methodological of the previous study by addressing sequence-based challenges, because previous study focused on improving ML models like Random Forest and XGBoost but in this study focused on temporal and spatial dependencies using LSTM with attention. LSTM with attention is the mechanisms has even higher computational complexity. Advanced feature engineering with spatial sequences, but the methodology might not suit all datasets. It is done by (Ma et al.; 2021).

In this study we create a hybrid model which leverages of both sequence handling and feature importance. And this study focusing on once particular city and used XGBoost to achieve high accuracy. Its fully focused on ensemble learning offering a unified framework for tackling diverse prediction challenges. But in these lacks feature engineering for advanced temporal or spatial analysis. This study is done (Ahtesham et al.; 2020).

This study aims to evaluate ANN based deep learning methods against classical methods like SVM, Random Forest and linear regression. ANN is performing well against all traditional models. In this study focused on all kind of metrics as well. But in this study exploring the ANN architectures with dense layer, showcasing their flexibility in handling the complex connectivity. Demonstrates ANN's potential for scaling up to larger datasets with more features. This study also focused on region but not detailed of correlation and feature engineering. Combine of previous study this study performed lack comparison with neural network. It is done by (Xu; 2021).

This study performed on stochastic gradient descent (SGD) and neural networks to improve prediction accuracy. Accuracy and precision metrics, along with graphical comparison of actual vs. predicted prices. In this research faces overfitting and suboptimal performance for high-priced houses due to limited training data. Focuses heavily on numerical features; no discussion of incorporating categorical or derived features. SGD-based training is computationally efficient but may oversimplify relationships for complex data. Research done by (Ni; 2022).

This study aims to use traditional machine learning, deep learning and hybrid model for time-series forecasting. In this research mainly focusing on proposed of hybrid model and deep learning. In this they performed well on normalizing the data and more into preprocessing. In this analysis they used small data so feature diversity bit lacked. This study was done by (Nunna et al.; 2023).

This study aims to perform data augmentation techniques. Here they implemented XGBoost, LSTM, CNN, RNN with data augmentation techniques like clustering and box-cox transformation, but in this mostly focusing on techniques performs. LSTM outperformed other models, for this study they used smaller dataset and limit scalability, (Shen et al.; 2023).

This study motivates us to address the data limitations and instability by data augmentation and predictive of LSTM with augmented data. Achieves 30% error reduction in LSTM with data augmentation. These are highly effectiveness for improving model performance. This research is slightly different from other research, focusing on addressing small dataset issues expands applicability to resource-limited scenarios, this study as well facing tackling issues like nonlinearity and dataset limitations. It is done by (Guang and Zubao; 2023).

This study Focuses on leveraging deep learning models, specifically CNN, for capturing non-linear relationships in housing price predictions. Effectively utilizes CNN, which got high accuracy. But in this study Only compares CNN with BPNN, leaving out comparisons with advanced models like GRU or LSTM and Lack of exploration of other advanced models like GRU and LSTM. This research was done by (Zhan et al.; 2020).

In this research detailed about ANN, CNN, RNN and KNN model performance using house price prediction. Based on lower RMSE value to comparing with traditional algorithms such as random forest and gradient boosting. But ANN is performed with high values. But this study needed more data variables, this research did with less volume of data variables. This study was done by (Karthika et al.; 2024).

This research aims at developing a hybrid deep learning model (CNN+LSTM+FCL) for capturing non-linear relationships in housing data and improve prediction accuracy compared to traditional methods. Effectively integrates CNN for feature extraction, LSTM for temporal trends, and FCL for final predictions, reduced error in RMSE and MAE. At the same time Focuses only on CNN, LSTM, and FCL, excluding exploration of other advanced deep learning models, limited analysis of external variables. But still more concentrated needed in data preprocessing, it relies heavily on structured and comprehensive datasets; missing or noisy data can significantly degrade the performance. In this study discussed more about implementing deep learning in effective manner, this research done by (Sakri et al.; 2024).

This research motivates us to do accurate price prediction assist with valuable variables to using of advanced deep earning model of CNN and ensemble methods of XGBoost, Random Forest, AdaBoost to enhance accuracy. They accurately combining of structed and unstructured data like housing attributes and images. So, they proposed a stacking ensemble approach. In this research CNN achieved more accuracy compared with other models. Balancing of complexity of model is one disadvantage on this research and its performing second-hand single houses so images and details are not good and needed to perform more in Data preprocessing this complete study done by (Srirutchataboon et al.; 2021).

This research aims to addresses the lack of comprehensive comparison among popular machine learning and deep learning methods for housing price prediction, here ML

models of SVM, Baysian, Linear Regression, DNN are used and SVM performed well in prediction. In this data preprocessing worked very well and ensured data quality as well. In this analysis linear regression and DNN are performed worst because of its smaller data. This research was done by (Chen et al.; 2021).

In this research aims to predicting house prices implementing of DNN in detail. In this they well detailed in data preprocessing to normalize the numerical features to ensure consistent scaling and detailed EDA. Well, detailed of hyper parameter optimization so got more remarkable accuracy and considered multiple evaluation metrices. As they used not a larger dataset and using multiple optimization techniques increases training overhead, which might not be practical for applications requiring quick predictions, this research done by (Vijava et al.; 2023).

2.3 Discussing House Price Prediction on Machine Learning

This research aims to use predict house prices in Malaysia. They applied ML and ANN as well. Here, AI and ML techniques enhance predictive capabilities. And in this analysis, They faced issue is limited data updates, not having a unique property factor. In this research they missed collecting proper data and not into detailed data preprocessing. This research was done by (Chee Kin et al.; 2022).

This research mainly focusing based on location on house prices fluctuation, in this research they implemented all regression model, but the ensample model of Random Forest and XGBoost performed well based on evaluation matrices of MAE. In this research they not properly handled categorical data and showing not into proper data preprocessing these all done by (Hamami and Dahlan; 2024).

2.4 Related work Outcomes

The review of all the above literature highlighted the various approaches for methodology, choosing larger data for advanced deep learning techniques. Mainly, the challenges during implementation of deep learning techniques, gave more solutions and suggestions. At the same time, detailed the importance of data preprocessing for implementing models. All kinds of variables handling methods were explained in the works. These studies also emphasize the importance of visualization, how the visuals helped in making decisions. However, challenges persist, including dataset limitations, overfitting in deep learning models, and the need for advanced preprocessing techniques. And for detailed knowledge on implementing deep learning techniques and comparative Machine learning, deep learning is required. For my research, these studies provided a complete foundational framework for developing deep learning techniques. By optimization of models, extensive preprocessing, evaluation metrics. Finally, these studies more gave more insights on implementing deep learning techniques on Ireland House Price Prediction.

2.5 Research Gap and Research Niche

The proposed research question aims to explore how advanced neural networks like ANN, RNN, CNN, and LSTM, combined with better data preparation and regression models, can improve house price prediction at the similar time this research focused on country of Ireland house price prediction. It focuses on making these models work well with large

datasets, across different housing markets, and in real-time applications, addressing gaps in current research.

3 Methodology

The main motive of this research is to predict the Ireland house price, which is different from common data analysis approaches. This study looks at existing research to understand the following methods as shown in as shown in figure1. The methodology is designed specifically for the dataset used in this research and follows this systematic and structured methodology.

This research is particularly based on KDD. The project begins with knowledge from data. Predicting the Ireland house prices based on various features like location, year, etc. And the complete cycle of KDD started like data collection, data preprocessing, model implementation and model evaluation and results.

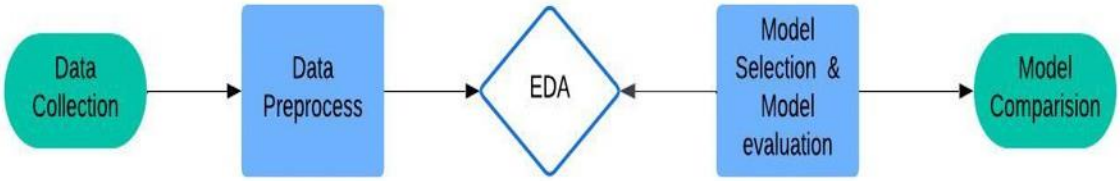

Figure 1: Flow of Methodology adopted to for our research study

3.1 Data Collection

Data collection is a major part of the research, because collection of data makes it more effective through research objective. The dataset used for this study is collected from Kaggle, a open-source data community called Property Price Register Ireland Dataset. It is publicly available data, and it maintains ethical standards. It contains main cities as well as important county data. This dataset is one of the main sources of the Ireland house prices details. This dataset contains data from 2010 to 2021 with 476745 rows $\times$ 9 columns of data as a CSV file.

3.2 Data Understanding and Data cleaning

Collected detailed Ireland house price data columns has its own dataset format, datatypes, noises of data, null values, error values were found based on research objective. Once errors were found; the complete dataset is processed for preprocessing. In this step, all the data are arranged for further processing. Each variable was checked in a detailed manner, focusing mainly on price columns. And the identified null values were removed from the columns. After these steps, basic cleaning was started in detail to identify outliers in the Ireland house price data, those were removed in proper manner. Finally, the data quality was ensured for further process. All these steps very clearly explained in implementation part.

3.3 Data preprocessing

Once the interpretation of the data and cleared all errors based on data cleaning method data into preprocess method. In the preprocess part, data was completely prepared for implementation model. A column is changed to numerical values. In this part, data were normalized through feature engineering methods. All the columns and variables were normalized in this part. These methods were helpful in balancing the complete data-set. Working on complex price data, cleaning and normalization and balancing complete datasets majorly helped in further process of model selection and model implementation. the preprocessing step made sure that the complete dataset was ready for implementing models.

3.4 Data Exploration/ Visualization

In general data, exploration and visualization is the critical part in understanding the full dataset and identifying price patterns, seasonal trends based on year and areas, relation-ship between each variable. So, this part completely works on Exploratory Data Analysis (EDA) and typically involves summarizing the data and generated needful insights. In this stage, pre-processed data is fed into EDA, were all the results were explained through visuals as charts and plots. The research majorly focused on target variables. Sales price of house is the target variable in this research. The data is used to identify distribution of price in every county, yearly or monthly average prices ranges in past years, correlation of the numerical features and seasonal trends. These all steps are detailly explained in implementation part.

3.5 Data Modelling

In data modelling, for the further process of feature engineering, preprocessed data was used during the implementation model. Before implementing the model, data is split into test and train. So, implementation of ANN, CNN, RNN, LSTM models were preparing from here. Whatever models are implemented, performances were also calculated. Based on target variables and test- train data, deep learning techniques were applied and fol-lowed by performance evaluation as well. For improved performance, Hyperparameter was applied along with early stopping, random and hyper tuning. All the techniques were implemented to identify model performances.

3.6 Model Evaluation and Predictions

Implementation of ANN, CNN, RNN, LSTM models were evaluated on the training and validation loss, which was not done in the training part. These are detailly explained along with evaluation of performance of the models. These evaluations followed some statistical metrics, based on those values to finalize the model performance. The metrices are RMSE, MSE, MAE which are determined to find the accuracy of the research. Hyper tuning assisted to increased models' performances. Once evaluation stage is completed, the high-performance model is used to compare predicted values and actual values which will be explained in the final part.

4 Design Specification

This research aims to predict the Ireland house prices across the complete country, so dataset was collected from a valuable source. The dataset is mixed with all counties and cities. The data used for analysis was in CSV format. The dataset imported into python to start the analysis. Beginning with data cleaning to extract data in detail, detailed cleaning process was undergone, missing values, columns transformation, outliers identified and removed, based on research objective. Further process of normalization also applied for model implementation. By Feature engineering based on research perspective, categorical variables are handled. Exploratory Data Analysis was applied on prepared data. In all the perspective, EDA was applied to visualize the data for better understanding. Once the data was prepared, the data is ready for implementation models based on research objective. After splitting the data as test and train, all four deep learning models are implemented. All hyper parameters and hyper tuning were used for improving model performance. Followed by evaluating model based on RMSE, MAE, MSE values. Finally, the models were compared based on evaluation metrics as shown in figure2.

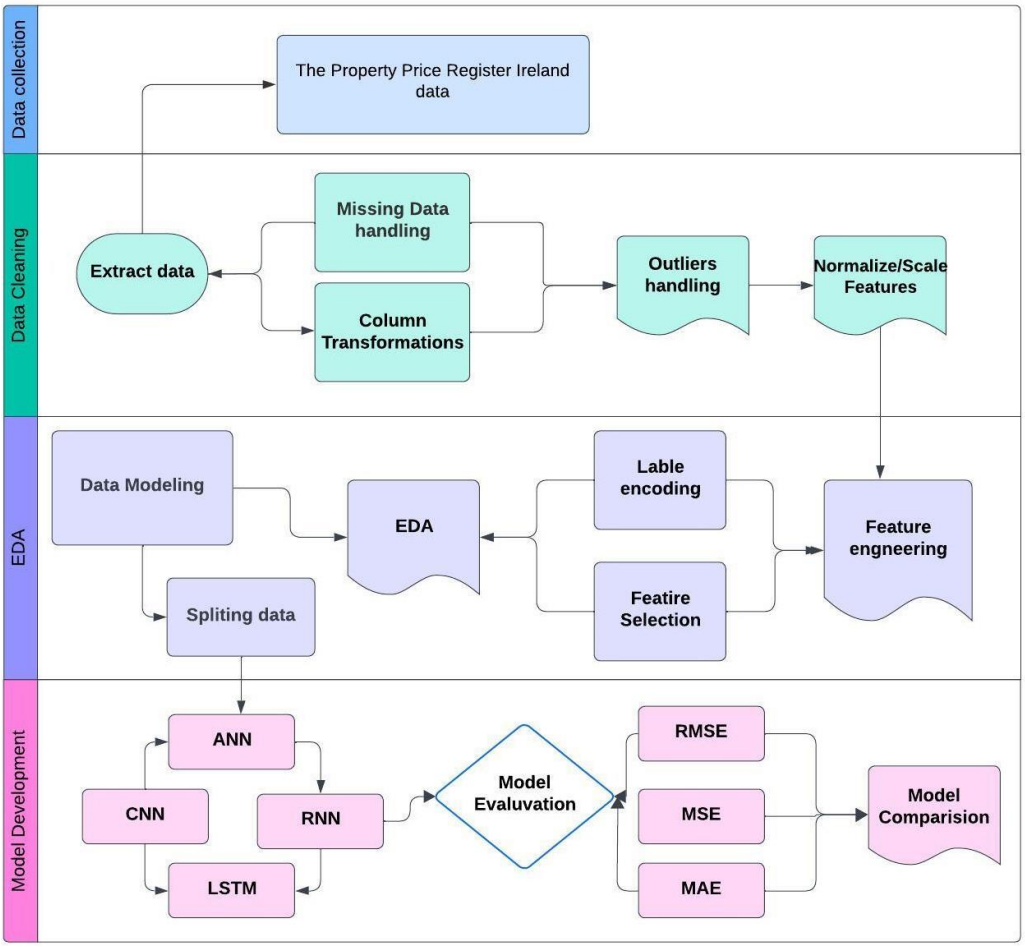

Figure 2: Design architecture of the research

5 Implementation

The section briefs step-by step implementation of deep learning model along with the modifications made during the implementation process. It includes data cleaning, data preprocessing, EDA, feature selections, applying of models and evaluation. The overview of methodology and implementation is provided below.

5.1 Data Collection

The Property Price Register Ireland dataset is collected from Kaggle, an open-source data community which covers the period from 2010 to 2021. The dataset contains 476745 records with multiple valid columns. Due to the huge volume of information, data preparation stage faced several challenges during this research. Figure3 shows the processed information data.

```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 476745 entries, 0 to 476744
Data columns (total 9 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   SALE_DATE        476745 non-null  datetime64[ns]
 1   COUNTY           476745 non-null  object
 2   SALE_PRICE       476745 non-null  float64
 3   IF_MARKET_PRICE  476745 non-null  int64
 4   IF_VAT_EXCLUDED  476745 non-null  int64
 5   PROPERTY_DESC    476745 non-null  object
 6   DAY              476745 non-null  int32
 7   MONTH            476745 non-null  int32
 8   YEAR             476745 non-null  int32
```

Figure 3: Information of preprocessed data

5.2 Data Cleaning

As mentioned earlier during the methodology section, the preprocessing of data including data cleaning is clearly detailed in this section. This stage involves identification and removal of missing values, finding outliers and a suitable method to remove outliers from the dataset. These processes are more helpful in improving the predictions.

- **Handling of Missing values:** From the data postal code and property size description, the missing values columns are unnecessary for the research objective, hence these columns were completely removed from the dataset. These two columns contained approximately 50% of missing values, which were removed during the research.
- **Handling of Outliers:** Further, Interquartile Range (IQR) method was used in identifying and removing outliers from the dataset (2023). In this method, lower bound and upper bound were defined to find the range within which most of the data should lie, assuming a normal distribution. Applying this method, 5340 outliers were found from the data and those were removed in proper manner as shown in figure4. Sales price was set as the target variable during the research.

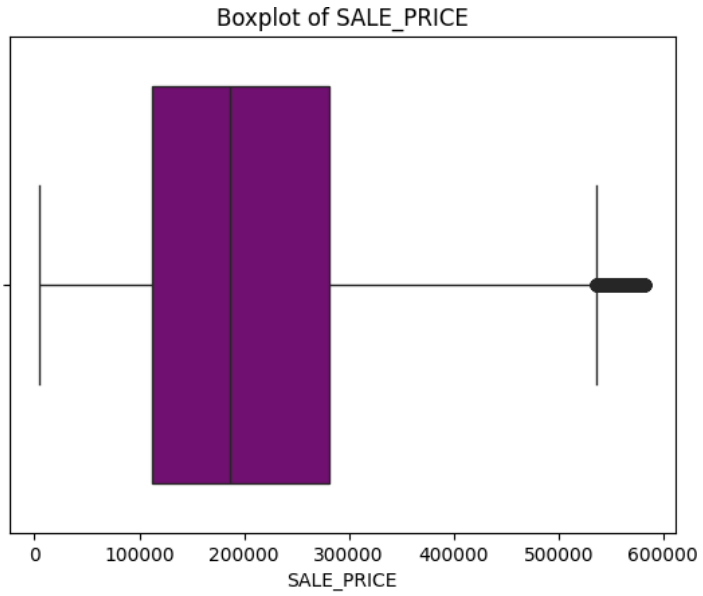

Figure 4: Outliers Visualization

5.3 Data Preprocessing and Feature Engineering

In this preprocessing stage, the cleaned data from data cleaning is used. In this process, data is gone through detailed analysis based on the research objective. So, column transformations are performed on this data, applying some feature engineering techniques for preparing data into implementing models.

5.3.1 Columns Transformations:

In sales data, columns are separated into date, month, year as three extra columns and sales data columns are dropped as well, because the research is focused on year wise change in Ireland house price prediction. So, these created columns are well used in this analysis.

5.3.2 Details of columns:

Once done with preprocessing, check all the columns data types and column names and verify if any changes are needed. This stage mainly focusses on categorical columns.

5.3.3 Handling unnecessary variables:

In the process, from the description column, nearly 35 variables are meaningless for this research, so these unwanted values were removed from the data.

5.3.4 Label Encoding:

Based on the target variable, one price category column is created in this data. This label-encoding is one of the most effective techniques for converting categorical variables into numerical values. Based on the highest and lowest price to categorize the data.

5.3.5 Encoding of Categorical Variables:

Using One-hot encoding, all categorical variables were transformed into numeric format, for ensuring the data to be suitable for modeling.

5.3.6 Feature Scaling:

Standard scaler method was used to standardize the numeric features by scaling into distribution to ensure that they contribute equally to the implementation of deep learning models. This step is a critical preprocessing stage for many models but is more relevant.

5.4 Exploratory Data Analysis

The exploratory data analysis was performed on Ireland house price dataset, providing important insights into trends, distribution, and relationships of key variables. The mentioned visuals impact the preprocessing methods. The primary goal of this analysis is to understand the house prices distributions, seasonal trends for the year and the relationship between year and county wise.

- **Trends of sales price:** Figure5 shows the complete details of sales price averages over the years. Initially from 2010 to 2013, the values were significantly dropping which can be seen from the graph. In those periods, the real estate markets were greatly decreased and demand for house prices also reduced. From 2013, there was an average sale price steady recovery, price range is mentioned over here. From 2016 to 2012, there was a rapid growth of sales phase in Ireland due to huge urbanization and housing demand.

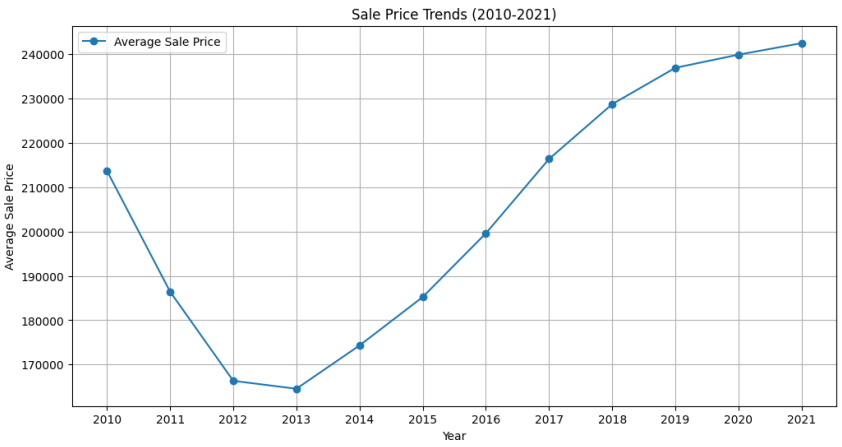

Figure 5: Trends of sales price over the year (2010 to 2021)

- **Distribution of price:** Figure6 shows that the density curve is smooth in histogram, providing a clear view of the general trend. The graph is positively skewed, with most houses with higher prices dominant, while fewer houses have significantly higher prices.

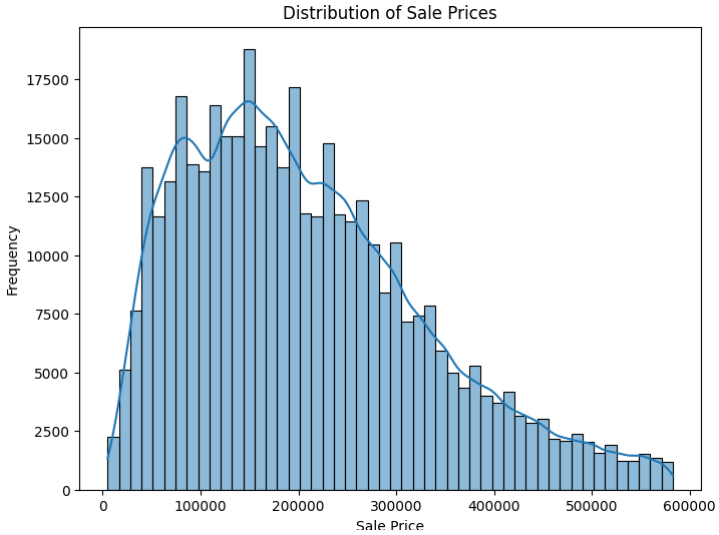

Figure 6: Sales Price distribution

- **Sale Prices by county:** Figure7 shows variation in house prices across the locations. Here, Dublin leads with the highest average sale price followed by Wicklow, Kildare, and Meath. Counties like Leitrim, Longford, and Roscommon have the lowest average sale prices. Counties like Cork, Galway, and Limerick are in the mid-range, indicating moderately high average sale prices. So, based on the geographic disparities prices are differing.

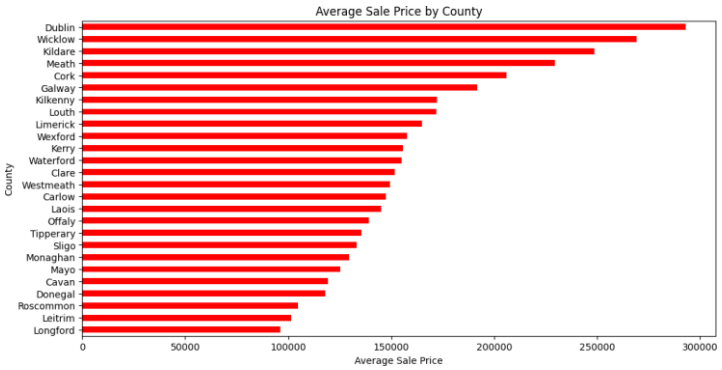

Figure 7: County-Wise Sale Price Analysis

- **Correlation:** The correlation analysis details the relationship between variables and highlights which features significantly impact the target variable. This correlation helps to detail data interpretation and insights. Figure8 shows correlation is more feasible.

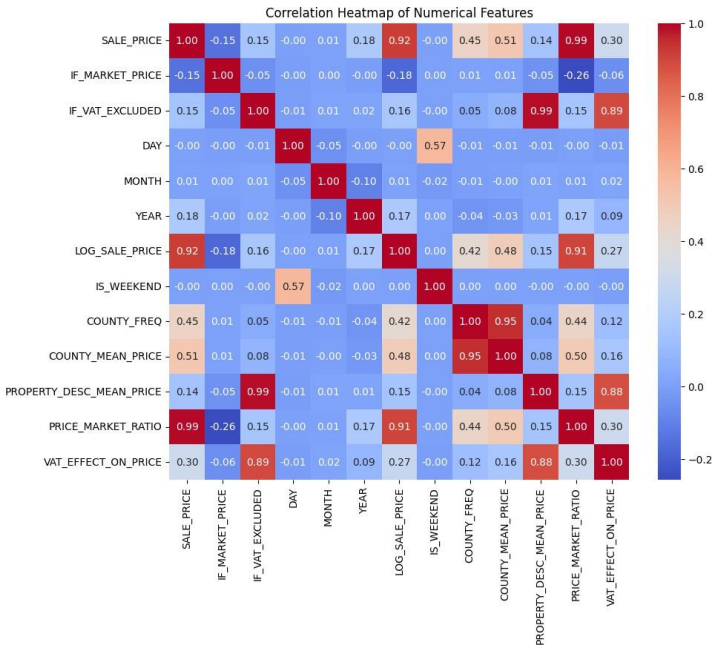

Figure 8: Correlation Analysis

- **Price Category:** Figure9 clearly explains average sales price based on their description i.e. Low, Moderate, High, Very High, Luxury. Properties are classified as high which exceeds 500,000, those are high-end properties on the market, along with moderate and low as well differencing. Pricing highlights clear market segmentation by property type.

Figure 9: Property Description Analysis

5.5 Model Development

In this section, Deep learning models such as Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory Networks (LSTMs) were implemented. These models were selected from the literature review approaches. The literature review of general house price prediction gave a clear idea of the above-mentioned deep learning models. The research gap and research niche provided a detailed idea undertaking this research using these four deep learning models. The implementation and performance of these four models are explained below.

- **Artificial Neural Networks (ANNs):** ANNs are designed to model complex relationships in data through layers of interconnected neurons. In this study, a three-layer ANN was employed using ReLU activation and dropout to prevent overfitting. The model was optimized with the Adam optimizer and evaluated using metrics such as RMSE, MAE, and MSE, achieving the best performance among all models.

- **Recurrent Neural Networks (RNNs):** RNNs are suited for sequential data by retaining infonnation across time steps. A simple RNN was implemented with ReLU activation and dropout, focusing on capturing temporal dependencies. While effective for short-term sequences, it underperformed compared to other models due to limited long-term memory handling.

- **Convolutional Neural Networks (CNNs):** CNNs, typ ically used for spatial data, were adapted for this study with convolutional layers and dropout to identify patterns in the dataset. Despite their robustness, CNNs showed limited performance for this regression task, as they are better suited for image-related data.

- **Long Short-Term Memory (LSTM):** LSTMs extend RNNs by incorporating memory cells to capture long term dependencies. This model leveraged dropout and ReLU activation to process time-series data effectively, performing better than RNN but slightly below ANN in terms of prediction accuracy.

In the above-mentioned implementations, hyper parameters are layer configuration of neurons and are adjusted based on models, dropout rates were adjusted, batch size , adam optimizer, Epoch and Early stopping were performed as well. Here, all models were run with epoch and early stopping, which lead to better performance.

6 Evaluation

This section provides an overview of evaluation metrics of the four techniques. The three metrics used in the evaluation of the four models are: RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and MSE (Mean squared Error). The four techniques evaluated are as follows.

- **Artificial Neural Network (ANN):** The ANN model performed the best, with minimal error metrics (MSE, RMSE, and MAE). Its predictions were highly accurate, making ANN the most suitable model for this regression task.

- **Recurrent Neural Network (RNN):** The RNN model shows a significant increase in error compared to ANN, it's higher RMSE and MAE values suggest that it struggles to make precise predictions, due to the lack of long-term memory capabilities inherent in RNNs.

- **Convolutional Neural Network (CNN):** The CNN model has the highest error metrics (MSE, RMSE, and MAE), performing the least compared to other models. CNNs are better suited for spatial data (like images) rather than sequential regression tasks, explaining its lower performance.

- **Long Short-Term Memory (LSTM):** The LSTM model performed better than RNN, benefiting from its ability to retain long-term dependencies in sequential data. Although its error metrics are higher than ANN's, they are significantly lower than RNN's, making LSTM a strong contender for this regression task.

Table 1 provides a detailed explanation of the result. Based on minimal error metrics Artificial Neural Network (ANN) performed well in comparison to other deep learning techniques, offering a more accurate and reliable model for predicting Ireland house prices. In general, ANN is the most effective and optimal choice for these kinds of regression problems.

Models	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
ANN	2510.298456	1682.450483	6301598.3393
RNN	11380.7728	9099.2021	129521990.2486
CNN	13297.9670	8534.6090	176835927.6497
LSTM	6181.7690	4060.1167	38214268.5791

Table 1: Evaluation Metrics

Actual Value	Predicted Value
130000.0	129859.4765
505000.0	510841.0000
248000.0	248782.8281
245000.0	244364.4531
236000.0	237058.1093

Table 2: Prices Actual s vs Predicted

Above mentioned Table 2 validates the actual price in comparison to Ireland predicted prices based on high performance model. Since Artificial Neural Network (ANN) performed well with lower error metrics, ANN model was used to predict the prices and compared with actual price. Both predicted and actual prices were consistently similar, i.e., predicted prices from Table 2 resemble the value in figure10. Figure10 shows the scatter plot of actual and predicted values under ANN techniques.

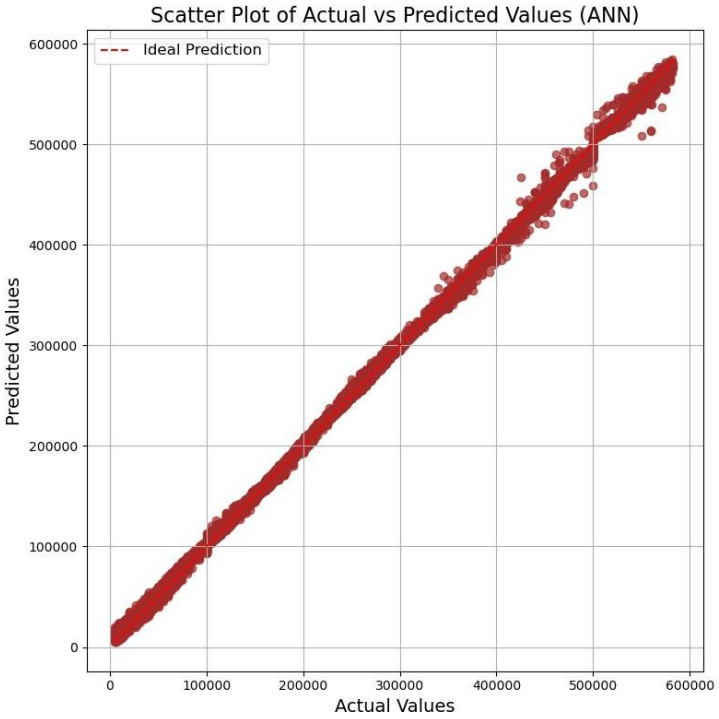

Figure 10: Actual values vs Predicted values

6.1 Discussions of result

This research implemented four advanced deep learning models of ANN, RNN, CNN, and LSTM to predict house prices in Ireland. Amongst the four models, ANN performed the best with the lower error metrics (RMSE: 2510.30, MAE: 1682.45, MSE: 6,301,598.34). LSTM showed moderate performance due to its capability in retaining long-term dependencies, while RNN and CNN struggled with sequential and regression tasks. Finally, ANN superior ability to model non-linear relationships, making it the most reliable predictor for Ireland's house prices. In this research, the above-mentioned techniques were effectively implemented. If an additional economic related vasriable could be added, the preprocessing could have been more effective with advanced hyper parameters. To improve the experiments, further variables need to be added like economic factors or house feature related columns. That would provide better ways for some advanced tuning. In accordance with previous research, advanced deep learning techniques for house price prediction were implemented, emphasizing the importance of ANN and preprocessing techniques in this prediction. This study extends earlier work by implementing multiple deep learning models, proving Artificial Neural Networks (ANN) dominance. Additionally, regional and temporal factors, as house prices change based on location and economic trends. Including details like county-wise economy, population growth, seasonal trends, and inflation rates can make predictions more accurate.

The ANN model performed better than other models because it could learn complex patterns in the house price pediction more effectively. Unlike RNN, which is works best for time-based sequences, or CNN, which is good for images, ANN is the well-suited for structured data like house prices. It used multiple hidden layers (128, 64, 32 neurons) with ReLU activation, allowing it to find important patterns in the data. Feature scaling and categorical encoding helped to ensure that all inputs were treated well, preventing

any one factor from dominating the predictions. The Adam optimizer made training faster and more accurate, while dropout regularization helped to prevent overfitting. Additionally, early stopping (after 150 epochs) ensured that training stopped when no further improvement was happened. These improvements are more helped ANN achieve much lower RMSE, MSE, and MAE values, created the best model for predicting Ireland's house prices compared to RNN, CNN, and LSTM, which struggled with capturing complex data relationships.

7 Conclusion and Future Work

The research processed advanced deep learning techniques such as ANN, RNN, CNN and LSTM. The primary goal was to identify the best-performing model and provide insights for Ireland housing price data from 2010 to 2021. Specifically, ANN emerged as the most effective model, with lower error metrics and created high reliability for Ireland house price prediction. Collecting this huge data of 476745 and increasing performance of model is the challenging part of this research. In the preprocessing stage, Exploratory data analysis was carried out with multiple variables like county, prices range and category wise changes. The study also highlighted important housing trends, such as regional price differences in past years as shown in figure5. The research successfully fulfilled the objectives. Future work can focus on incorporating regional and time-based factors to improve predictions, as house prices vary across locations and economic conditions. Adding the county-wise economic indicators, population growth, seasonal trends, and inflation rates can make the model more precise. ANN can also be optimized by combining it with LSTM to capture long-term trends and using separate models for different regions to improve location-specific accuracy. Additionally, integrating geospatial data, such as proximity to schools, transport, and amenities, can provide more realistic price estimates. These enhancements will make the model more reliable, region-aware, and time-sensitive, leading to better forecasting of Ireland's housing market trends.

References

Ahtesham, M., Bawany, N. Z. and Fatima, K. (2020). House price prediction using machine learning algorithm - the case of karachi city, pakistan, *2020 21st International Arab Conference on Information Technology (ACIT)*, pp. 1–5.

Chee Kin, C., Arabee Bin Abdul Salam, Z. and Batcha Nowshath, K. (2022). Machine learning based house price prediction model, *2022 International Conference on Edge Computing and Applications (ICECAA)*, pp. 1423–1426.

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)*, pp. 699–702.

Guang, W. and Zubao, S. (2023). Research on the application of integrated rg-lstm model in house price prediction, *2023 IEEE 5th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pp. 348–353.

Hamami, F. and Dahlan, I. A. (2024). Regression modeling for house price prediction in java island, *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, pp. 402–406.

Joseph, L., Reddy, E. H., Srinidhi, M., Saketh, P. V., Mohan, D. and Prasad, C. R. (2024). Predicting real-time house prices: A machine learning approach using xgboost algorithm, *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*, pp. 1–4.

Karthika, K., Balasubramanie, P., Dharshini, K., Shanmugapriya, P. and Ramya, T. E. (2024). Harnessing artificial neural network model to anticipate housing market rates, *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–8.

Ma, S., Jiang, G., Liu, J. and Wang, X. (2021). Research and application of second-hand housing price prediction model based on lstm, *2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, pp. 320–325.

Ni, Y. (2022). A housing price prediction method based on neural network, *2022 International Conference on Big Data, Information and Computer Network (BDICN)*, pp. 592–595.

Nunna, K. C., Zhou, Z. and Shakya, S. R. (2023). Time series forecasting of u.s. housing price index using machine and deep learning techniques, *2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pp. 1–6.

Saini, K., Ramvachan, Kashyap, P. and Yadav, P. (2023). Prediction of house pricing using machine learning, *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 131–135.

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)*, pp. 127–133.

Shen, Y., Wen, L., Yang, X. and Li, Z. (2023). Housing price prediction based on data augmentation, *2023 42nd Chinese Control Conference (CCC)*, pp. 8300–8305.

Srirutchataboon, G., Prasertthum, S., Chuangsuwanich, E., Pratanwanich, P. N. and Ratanamahatana, C. (2021). Stacking ensemble learning for housing price prediction: a case study in thailand, *2021 13th International Conference on Knowledge and Smart Technology (KST)*, pp. 73–77.

V, H., B, D. B., R, S., R, V. and D, R. (2024). A comparative study of house price prediction using machine learning and deep learning techniques, *2024 Third International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, pp. 1–6.

Ved, A. R. and Gupta, A. (2024). Extensive data preprocessing and exploratory data analysis for house price prediction using machine learning and deep learning with an introduction to automl technologies, *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, pp. 711–721.

Vijava, J., Kumar, P. S., Sorgile, M. and SV vasanth, M. S. (2023). Optimization techniques for deep learning based house price prediction, *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, pp. 202–207.

Xu, J. (2021). A novel deep neural network based method for house price prediction, *2021 International Conference of Social Computing and Digital Economy (ICSCDE)*, pp. 12–16.

Zhan, C., Wu, Z., Liu, Y., Xie, Z. and Chen, W. (2020). Housing prices prediction with deep learning: an application for the real estate market in taiwan, *2020 IEEE 18th International Conference on Industrial Informatics (INDIN)*, Vol. 1, pp. 719–724.