HOUSE PRICE PREDICTION IN THE UNITED KINGDOM (UK)

MSc. Project Report

Osinuga Timilehin Deborah

202236224

University of Hull

Supervisor: Oseikhuemen Davis Ojie

Abstract

Housing price prediction stands as a frequently explored subject with substantial consequences for informing stakeholders and real estate buyers. This study delves into the suitability of traditional machine learning, ensemble learning, and deep learning techniques for predicting real housing prices using UK house price paid data. The analysis reveals that in the context of forecasting time series data, deep learning, specifically with the use of ANN and LSTM models, performs exceptionally well. On the other hand, when predicting housing prices based on multiple influencing factors, ensemble learning, particularly with CatBoost, outperforms traditional machine learning, achieving an improved fit with an R^2 of 0.80, with key factors such as country, year, town/city, and property type impacting house prices. These findings emphasize the effectiveness of ensemble learning techniques, highlighting the robust performance of the LSTM model in forecasting house prices, with an exceptional accuracy of 96%, surpassing previous works.

Keywords: House Price Prediction, LSTM, CatBoost.

1 Introduction

A home surpasses its physical confines; it constitutes an invaluable resource offering more than just a place for shelter, security, and personal identity. It also carries substantial financial significance, especially in terms of its potential as a prudent investment. (J Richardson, 2018). The dynamics of the UK housing market are influenced by historical factors, economic shifts, and the evolving preferences of its residents. As individuals and entities engage with the complexities of the real estate landscape, the ability to anticipate and understand house price movements becomes paramount.

Lancaster's Characteristics Theory introduces the concept of housing as a bundle of characteristics, each contributing to its overall utility (Lancaster, 1966). This theory emphasizes that individuals make housing choices based on the specific attributes of a property, such as size, location, and amenities. Lancaster's approach provides insight into the nuanced preferences that influence housing demand and, consequently, prices.

Each year, increasing housing demand drives up prices, influenced by factors like location and demand. Stakeholders, including buyers and developers, seek to understand these influences for better decision-making. Between 1991 and 2014, there was a growth in London's population and household count, which raised housing demand. An average house price increased by 9% year on average between 1998 and 2014, largely due to the mismatch between the growth of the housing stock and demand (Marsden, 2015). House price prediction is crucial for homeowners, aiding investment decisions, and for policymakers to formulate effective housing policies. Predictive models provide proactive insights, empowering stakeholders in their decision-making processes.

Within the real estate market, conventional forecasting techniques have long been utilized to characterize existing market conditions (J Pallesen, 2018). Nevertheless, these approaches prove insufficient in delivering highly accurate assessments when confronted with vast datasets. The advent of machine learning, endowed with the capability to process extensive data sets and discern intricate trends, has positioned itself as a formidable tool for the prediction of house prices (Lantz, B., 2019). Employing techniques such as regression models, decision trees, and neural networks allows for the development of precise and adaptable predictive models, reshaping the landscape of house price forecasting.

1.1 Problem Statement

The current landscape of predicting real estate prices in the UK reveals a notable gap in the literature. While some studies, like Alen Vlahovlijak's (2022), have utilized regression models, they are constrained by the exploration of only linear and decision regression models and a dataset limited to 1995-2017. Despite the growing interest in applying machine learning to real estate, the UK house price paid data from the HM Land Registry (1995-2022) has yet to be extensively explored. Furthermore, the absence of ensemble learning techniques and the underutilization of time series forecasting, specifically employing Artificial Neural Networks and Long Short-Term Memory models, underscore a critical void in the research landscape.

1.2 Aim and Objective

The aim of this study is to assess the efficiency of deep learning, machine learning, ensemble techniques in accurately predicting housing prices in the UK.

- Collect housing price data in the UK from 1995 to 2022, specifying property type, location (town/city, postcode), and market trends. Clean and preprocess the data for effective modeling.
- Identify the pivotal features that have a substantial impact on housing prices in the United Kingdom.
- Examine diverse regression models to assess their effectiveness in predicting house prices.
- Compare and evaluate a range of traditional machine learning models, ensemble learning techniques, and deep learning models to identify the most precise and dependable model for forecasting house prices in the UK.
- Utilize ANN and LSTM to forecast house prices, examining their potential for accurate predictions in the context of the UK housing market.

2 Literature Review

The rise of advanced machine learning techniques, demonstrated by the powerful XGBoost algorithm, has ushered in a transformative era for house price predictions. This evolution is vividly demonstrated in the research titled "Assessing the Worth of Urban Green Spaces," (Joe Davies et al, 2019), which showcases substantial improvements of 84% in accuracy. This underscores its potential for precise and insightful predictions in real estate, highlighting its versatility and effectiveness in capturing complex relationships within datasets. The research not only contributes to more accurate property valuations but also signifies the broader impact of advanced machine learning techniques in reshaping the landscape of real estate analysis.

Boyapati Sai et al. (2022), explored the application of various machine learning algorithms and ensemble methods, including KNN, SVM, Linear Regression, XGBoost, Gradient Boosting Regressor, CatBoost, AdaBoost, and Random Forest Regression, is rigorously examined for house price prediction. The findings highlight the distinct advantage of ensemble learning over traditional models, with CatBoost emerging as the top performer with R^2 of 80% underscoring its superiority in predictive accuracy.

(Lee et al., 2022) introduced Long Short-Term Memory (LSTM) for real estate price prediction, achieving a solid 76% prediction similarity rate. The research not only emphasizes the importance of using different modeling techniques but also highlights how LSTM is crucial for analyzing real estate time series data. Despite their successes, the authors acknowledge limitations in a practical way and suggest getting more data to improve predictions beyond 76%, pointing out the lack of comprehensive real estate-related data. The idea of using public Big Data underscores the importance of having strong datasets to improve and advance predictive models.

Shi's (2023) recent study conducts a comprehensive assessment of machine learning (LightGBM) and deep learning (LSTM) models in predicting house prices. Notably, in the specific task of pure time series prediction based on historical housing prices, the LSTM model demonstrates an exceptional R-squared value, achieving an accuracy of 91%. This emphasizes the promising potential of deep learning techniques in capturing intricate patterns within housing data.

Alen Vlahovljak's (2022) investigation into Multivariable Linear Regression and Decision Tree for real estate price prediction, utilizing a dataset from the UK's HM Land Registry spanning from 1995 to 2017, brings attention to the complexities of real estate data. Despite cross validation, Multivariable Linear Regression struggles with consistently low accuracy scores around 42.57%. In contrast, the Decision Tree regressor significantly outperforms, achieving an average score of 75.91%. This success is attributed to the Decision Tree's ability to handle non-linear relationships and insensitivity to colinearity, addressing limitations observed in Linear Regression. The study not only underscore the importance of choosing models that can effectively navigate the complexity of real estate data but also emphasizes their essence in achieving accurate predictions crucial for informed decision-making in the real estate market.

3 Methodology

This study employs machine learning, ensemble learning, and deep learning techniques to predict housing prices in the UK. The illustrated approach for tackling the challenge of predicting house prices is presented in Figure 1.1. The following subsections will provide in-depth insights into each component depicted in the diagram, contributing to a comprehensive understanding of the methodology utilized in this study.

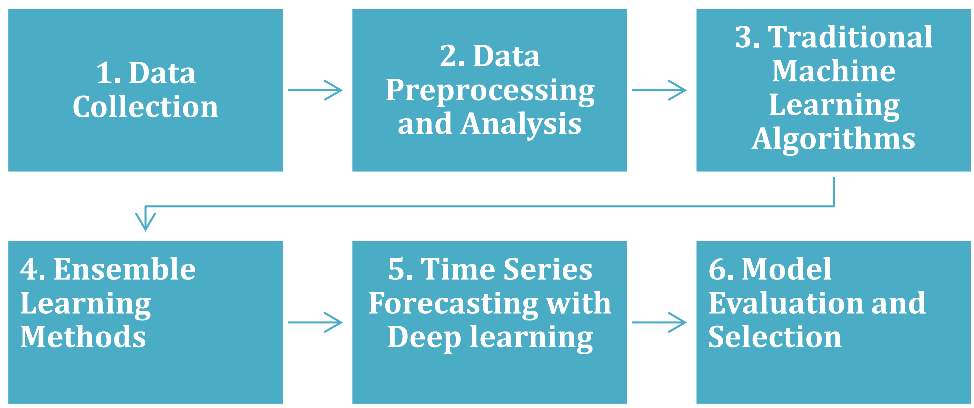

Fig 1. Proposed Method

3.1 Data Collection

The dataset employed in this study comprises comprehensive records of individual property sales in England and Wales dating back to 1995. Each record in the dataset includes a detailed property description along with its corresponding market value. The dataset has been sourced from the (HM Land Registry). It is made available under the Open Government License 3.0. Encompassing transaction records from 1995 to 2022, the dataset provides a wealth of information for analysis. The dataset comprises 28,275,940 records across 16 columns.

3.2 Data Preprocessing and Analysis

The analysis dataset was categorized into two groups: time series data and multi-factor analysis data. For the task of predicting housing prices using deep learning techniques in a time series context, the UK housing dataset was utilized, extracting key features such as the 'Date of Transfer' and 'Price.' The 'Date of Transfer' data was transformed into a time series format, and average house prices were calculated and used to enable time series analysis. In instances where inaccuracies in monthly average prices for specific years were identified, errors were rectified by incorporating precise information from the UK house price index to ensure the accuracy and clarity of our results.

After completing the data cleaning process, a systematic approach was employed to split the entire dataset into two distinct sets. To ensure a comprehensive evaluation of the model's performance, 70% of the data was assigned to the training set, covering the historical period from 1995 to 2014. The remaining 30% was designated for the test set, representing the more recent market dynamics spanning from 2015 to 2022. This segmentation enables the model to learn patterns and trends from past data, allowing it to capture diverse market conditions and trends over time. The approach takes into account significant events, such as the 2008-2009 financial crisis and subsequent recession, followed by a recovery phase in 2014 as highlighted by Lucian Cook (2017).

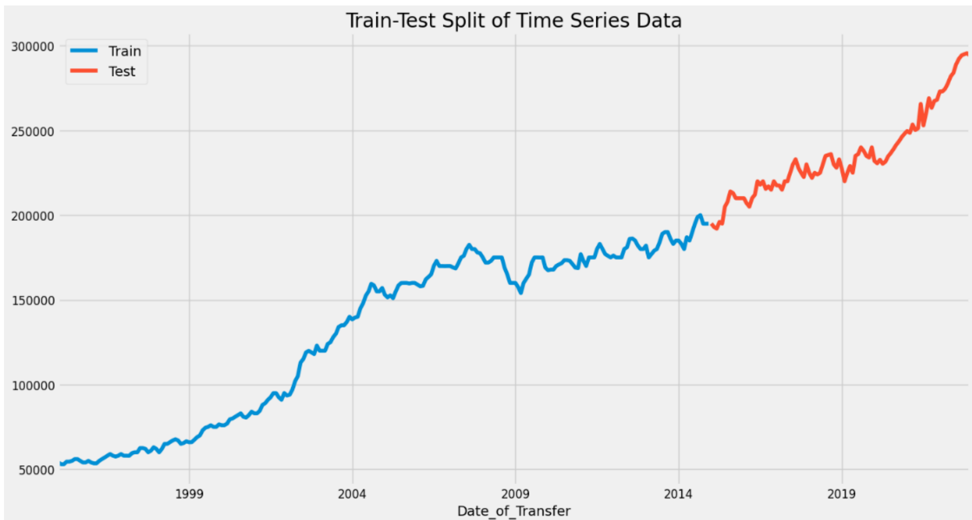

Fig 2. Train-Test Split of Time Series Data.

In the thorough data preprocessing for predicting house prices using multiple factors, the initial steps involved carefully addressing missing values by eliminating null entries in 'street,' 'PAON,' and 'postcode.' The temporal aspect of the dataset was improved by reformatting the 'Date of Transfer' and adding new 'year' and 'month' columns for more detailed analysis. To streamline the dataset, unnecessary columns like 'duration,' 'PPD category type,' 'record-status-monthly file only,' 'locality,' and 'SAON' were dropped. Refinement of the 'property type' feature included fine-tuning categories, particularly addressing the 'Other' category by excluding records with 'O' (Other) in the 'property type' to ensure analysis relevance.

Categorical features such as 'District,' 'County,' 'old/new,' 'Town/City,' 'property type,' and 'postcode' underwent one-hot encoding, creating new columns and removing the original categorical columns for better model compatibility. Dealing with data skewness became crucial, as the 'Price' variable showed a positive skewness of 42.03, indicating a longer right tail in its distribution. To address this, a log transformation was applied to normalize the data and enhance regression model performance. Outliers in the 'Price' variable were replaced with the mean value to mitigate the impact of extreme values. Subsequent normalization techniques were applied to the transformed data. For computational efficiency, random sampling (10% of the data) was used, and the dataset was split into training (70%) and testing (30%) sets. Various regression models were then trained on the preprocessed data, providing a solid foundation for robust predictions based on multiple factors.

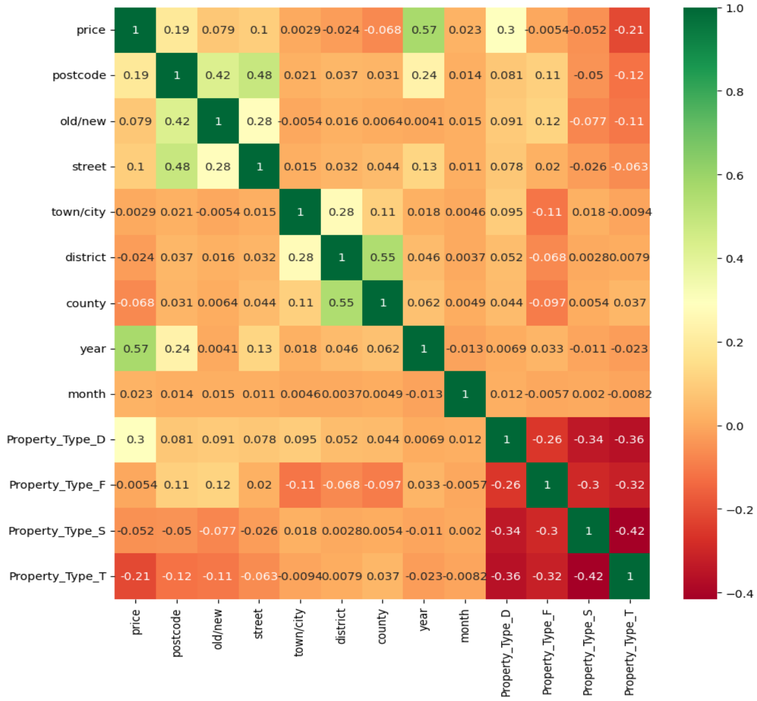

Fig 3. Heatmap with correlations

The price displayed a moderate correlation of 0.57 with the year, and the property type (Detached) also demonstrated a positive correlation of 0.3 with the price, indicating their influence on house prices.

3.3 Traditional Machine Learning Techniques

1. Linear Regression

Linear regression involves modeling the relationship between a dependent variable and independent variables by minimizing the sum of squared differences between observed and predicted values (Kaya, 2013).

2. Random Forest Regression

Random Forests function by employing an ensemble of decision trees that collaborate to make predictions. The ultimate output of the model is derived from the average of predictions generated by each individual tree. This approach reduces vulnerability to overfitting, enhances flexibility, and enables the modeling of nonlinear relationships between variables (Segal,2004).

3. Decision Tree Regression

Decision tree models can be applied to datasets that contain both category and numerical variables. They are particularly good at catching interactions that are not linear between the target variable and these attributes. Decision trees are also intuitive for understanding data since they show some similarities to human thought processes (Rathore, 2016).

4. K-Nearest Neighbors

K- nearest neighbors is a straightforward machine learning model. When faced with a new sample, it looks at the labels (the "k" closest ones) and assigns a label to the new sample based on what's most common among those neighbors (Yunsheng et al, 2017).

3.4 Ensemble Learning Techniques

1. **Cat Boost**: Cat Boost is a popular gradient boosting framework designed to effectively handle categorical features and big data in machine learning tasks. It stands out for its efficient treatment of categorical variables and incorporates various features that contribute to its robust performance (JT Hancock, 2020).
2. **XGBoost**: XGBoost, an acronym for eXtreme Gradient Boosting, is a particularly effective and optimized gradient boosting method. It performs better than conventional gradient boosting techniques (Avanijaa J, 2021).
3. **Stacking**: Stacking is an ensemble learning strategy wherein a meta-learner, often called a blender, is used to combine the predictions of several different models that have been trained separately to improve overall predictive performance. Stacking captures non-linear relationships that a single linear regression model might overlook. (L Breiman, 1996).
4. **Bagging**: The idea behind bagging is to combine the predictions of several instances of the same learning algorithm that have been trained on diverse portions of the training data. Bagging reduces model variance, improving robustness against overfitting. It enhances predictive accuracy, especially for unstable models. (Q Sun, 2012).

3.5 Time Series Forecasting with Deep Learning methods (ANN & LSTM)

Artificial Neural Network: ANN involves using neural network architecture to capture and learn patterns from historical time series data in order to make predictions. Artificial Neural Networks are a category of machine learning models designed with inspiration from the structure and functioning of the human brain, as discussed by (Guillod et al 2020).

Long Short Memory Network: LSTM is a type of recurrent neural network designed for recognizing long-range dependencies in sequential data. It excels in tasks like time series forecasting and can handle large datasets effectively, making it valuable for house price forecasting.

3.6 Model Evaluation

Model evaluation involves assessing and quantifying a model's performance to understand its ability to generalize to new data. To prevent overfitting and ensure robust assessment, all models underwent k-fold cross-validation. Various metrics, such as RMSE and R Squared value, were employed to measure accuracy and provide a comprehensive evaluation of the model's effectiveness.

1. R-Squared Value: R-squared is a metric used to quantify how well the independent variables account for the dependent variable's variability. It has a value between 0 and 1, with 1 denoting complete model explanation of the target variable's variability.

2. Root Mean Squared Error: RMSE, is a metric in statistics and machine learning that gauge’s predictive model accuracy. It computes the square root of the average squared differences between predicted and actual values, with lower RMSE values indicating superior model performance.

4 Experimental Results

Various techniques were employed to forecast house prices, comparing the efficacy of time series forecasting with predictions based on multiple factors. The results indicated that deep learning techniques, specifically Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM), outperformed other methods in forecasting house prices using time series data. In contrast, for predictions based on multiple factors, ensemble learning methods such as stacking, bagging, XGBoost, and CatBoost demonstrated superior performance compared to traditional approaches like Linear Regression, K-Nearest Neighbors (KNN), Decision Tree Regression, and Random Forest Regression. The detailed analyses of the findings are presented in Section 4.1 and 4.2, providing insights into the most effective models for predicting house prices in both time series and multiple factor contexts. The summarized results are presented in tables 1 and 2 and the performance comparison of all models is illustrated in fig 4.

Table 1: Performance For all Algorithms across folds

Folds	**Linear Regression**		**Decision Tree**		**Random Forest**		**KNN**		**XGBoost**	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
k1	0.446	0.573	0.656	0.451	0.775	0.365	0.567	0.506	0.785	0.356
k2	0.447	0.572	0.656	0.451	0.778	0.362	0.569	0.506	0.786	0.356
k3	0.446	0.573	0.655	0.452	0.777	0.366	0.568	0.505	0.786	0.355
k4	0.446	0.572	0.654	0.452	0.781	0.361	0.568	0.505	0.787	0.355
k5	0.445	0.573	0.655	0.452	0.778	0.362	0.566	0.506	0.786	0.356
Mean	0.45	0.57	0.66	0.45	0.78	0.36	0.57	0.51	0.79	0.36
Std	0.0008	0.0004	0.0005	0.0003	0.0013	0.0013	0.0008	0.0002	0.0005	0.0005

Table 2 Contd: Performance for all Algorithms across folds

Folds	**CatBoost**		**Bagging**		**Stacking**		**ANN**		**LSTM**	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
k1	0.797	0.35	0.767	0.393	0.777	0.364	0.976	0.03	0.923	0.049
k2	0.796	0.35	0.767	0.392	0.778	0.363	**0.98**	0.038	**0.989**	0.02
k3	0.796	0.349	0.761	0.396	0.772	0.362	0.844	0.032	0.923	0.053
k4	0.796	0.35	0.767	0.393	0.777	0.367	0.909	0.061	0.969	0.026
k5	0.796	0.351	0.765	0.393	0.777	0.364	0.972	0.039	0.902	0.013
Mean	0.80	0.35	0.77	0.39	0.78	0.36	0.94	0.04	0.96	0.038
Std	0.0004	0.0002	0.0022	0.0013	0.0022	0.0017	0.053	0.011	0.037	0.013

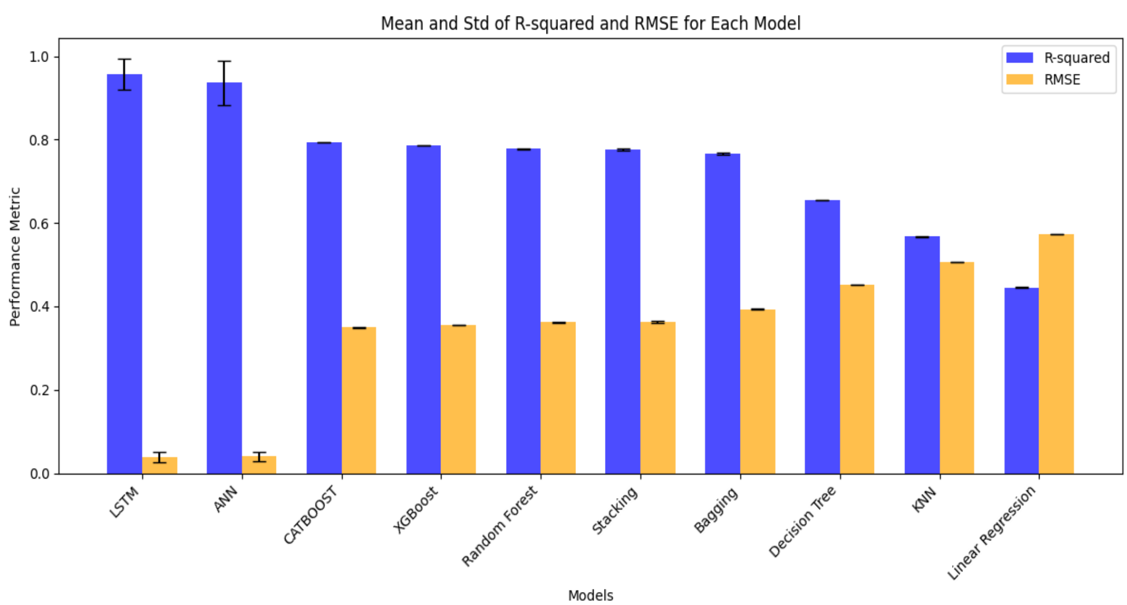

Fig 4. Performance Comparison: Mean and Std of R-squared and RMSE for Traditional, Ensemble, and Deep Learning Models.

4. 1 Forecasting based on Time Series

4.1.1 LSTM

The LSTM model, trained with 5-fold cross-validation over 100 epochs using the Adam optimizer, exhibited exceptional performance in time series forecasting. With a mean R^2 of 0.96 and a mean RMSE of 0.038, the model demonstrated high accuracy and stability. The architecture featured a single LSTM layer with 7 units, utilizing ReLU activation and lecun_uniform kernel initialization. Employing Mean Squared Error as the loss function, the model implemented effective early stopping (patience: 2) to prevent overfitting. Low standard deviations for R^2 (0.037) and RMSE (0.013) underscored the model's consistent performance across different folds. The model was tested on an independent dataset known as the test data, and the outcomes are visually illustrated in Figure 5.

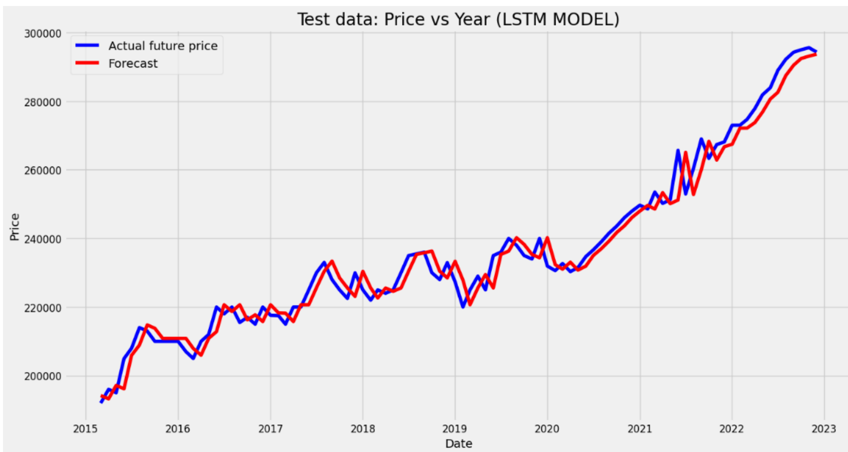

Fig 5. Test data on LSTM

The presented illustration highlights the excellent performance of the test data, showcasing a remarkable alignment between the forecast and the real price, surpassing the performance of the ANN model. The model effectively predicts the actual future prices spanning from 2015 to 2022, signifying the reliability and accuracy of the LSTM model in forecasting prices based on historical data.

4.2 Predicting based on Multiple Factors

4.2.1 CATBOOST

Following a comprehensive 5-fold cross-validation analysis, ensemble learning techniques, particularly CatBoost, demonstrated superior performance in predicting house prices compared to traditional methods. CatBoost exhibited a remarkable mean R^2 of 0.80 and an impressively low mean RMSE of 0.35, emphasizing its consistent and accurate predictions across different folds. The associated small standard deviations (Std R^2: 0.0004, Std RMSE: 0.0002) further underscore the stability of CatBoost's performance, indicating minimal variability in predictive accuracy during the cross-validation process. The model underwent testing on an independent dataset referred to as the test data, with the results visually depicted in Figure 6.

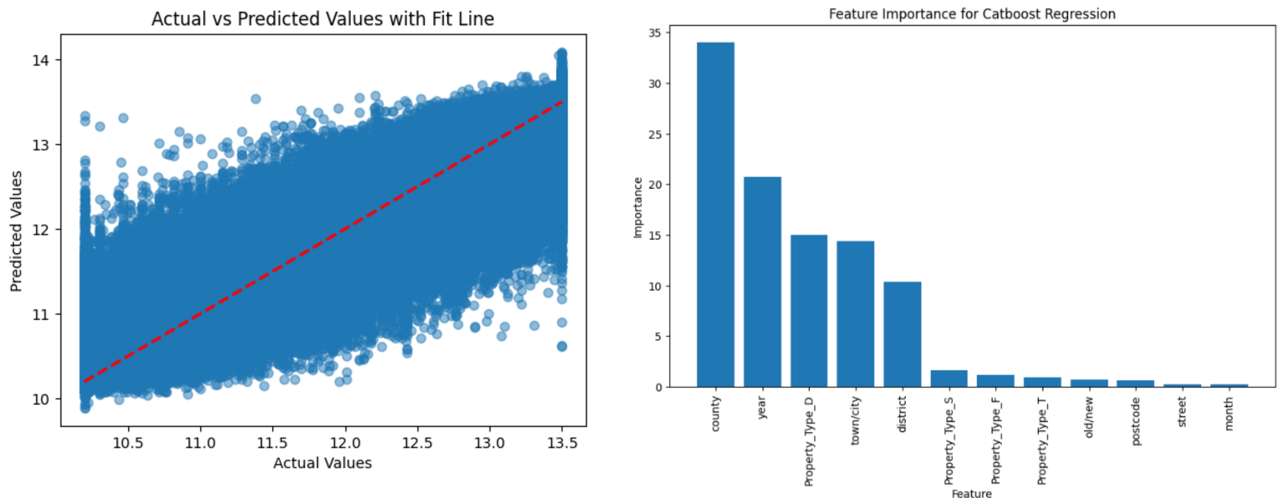

Fig 6. Test data: Actual vs Predicted Values and feature importance in CatBoost Regression

The visual representation showcases a robust correlation between predicted and actual house prices, underscoring the dependability of CatBoost in predicting house prices using diverse factors. Country stands out as the most impactful feature, with the year, property type-D, town/city, and district also play significant roles in predicting house prices.

5 Discussion

The analysis of UK house price paid data encompassed two primary aspects: predicting house prices through the consideration of multiple factors and forecasting house prices using time series analysis. Various models, including LSTM, ANN, Linear Regression, CatBoost, XGBoost, decision tree, and random forest, were compared to determine the most precise and consistent model for predicting house prices. The analysis revealed that LSTM emerged as more effective in predicting UK house prices.

The findings presented in the results section align with and build upon insights derived from the literature on traditional machine learning, deep learning, and ensemble learning. The study emphasizes the crucial role of forecasting real estate data through time series analysis, consistent with existing literature. The superior performance of the LSTM model compared to traditional approaches underscores the significance of leveraging deep learning for accurate house price prediction, as suggested by Shi (2023).

Consistently, the results highlight the effectiveness of CatBoost in handling categorical features and big data. This aligns with existing literature acknowledging CatBoost's ability to manage complex relationships within data, surpassing traditional models (Boyapati Sai et al., 2022). Despite dealing with large datasets and complex data, CatBoost's superior performance across five folds suggests its utility in handling big data and its value in real estate prediction tasks.

Furthermore, the study sheds light on XGBoost's ability to predict real estate prices effectively. The strong performance of XGBoost aligns with literature emphasizing the potential of advanced machine learning over traditional approaches (Joe Davies et al., 2019). Despite the complex relationships within the data, XGBoost significantly improved the accuracy of predicting house prices, almost comparable to CatBoost, establishing it as a reliable model for such predictions.

The study addresses challenges highlighted in the literature by Alen Vlahovlijak (2022), Among traditional models, Linear Regression exhibited poor performance, consistent with the literature, indicating its unsuitability for handling complex and nonlinear relationships for the UK real estate data. An improvement over traditional approaches was observed in Decision Tree, Random Forest, and KNN, emphasizing their effectiveness in handling nonlinear relationships within real estate data. Additionally, more advanced machine learning models like bagging and stacking were employed to enhance the accuracy of the traditional models.

6 Conclusion and Future work

The prediction of real estate prices involved an integration of Traditional Machine Learning, Ensemble Learning, and Deep Learning techniques, utilizing a dataset that was divided into two distinct categories: Time Series and Multiple Factors. Notably, within the exclusive Time Series category, the LSTM model outperformed traditional Machine Learning methods and also surpassed Ensemble Learning approaches. Conversely, in the category that incorporated multiple factors, CatBoost demonstrated superior performance of 0.80%. key factors like country, year, property type and town/city can guide buyers and investors in identifying opportunities and risks associated with a particular property or investment strategy.

The LSTM model emerged as a robust and reliable predictor, achieving an impressive accuracy level of 96%. Equipped with these insights, stakeholders can anticipate market trends and strategically address potential financial risks. In future research endeavors, there is merit in exploring and refining the integration of ensemble techniques with the Gated Recurrent Unit (GRU) model. Investigating the robustness of the GRU model across diverse datasets and geographical locations could yield valuable insights into its generalizability and applicability.

7 Degree of Difficulty

Predicting house prices using real estate data in the UK presented numerous challenges, adding to the complexity of the project. The large size of the dataset necessitated the utilization of Google Colab pro+, an A100 GPU, and a substantial 84 GB of RAM for effective data management and model training without encountering crashes. To tackle the vastness of the dataset, a sequential model-building approach was adopted due to computational limitations. This involved constructing each model sequentially, extending the overall project timeline and requiring meticulous planning to optimize resource utilization. Without random sampling, model choices are limited to only two regression models. Recognizing the significance of this limitation, the integration of random sampling became crucial. This strategic addition played a pivotal role in maximizing resources, improving computational efficiency, and striking a balance between the need for accurate and unbiased results. Furthermore, it facilitated the exploration of a more diverse range of regression models. In forecasting house prices, the deep learning models demonstrated excellent performance, efficiently handling the large dataset and minimizing training time, making them well-suited for time series tasks. The decision to incorporate traditional machine learning, ensemble learning, and deep learning proved effective in overcoming challenges and achieving accurate predictions of house prices.

References

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, *82*(1), pp.34-55.

Lancaster, K., 1971. Consumer demand: A new approach.

Granna, J., Brunauer, W. and Lang, S., 2022. *Proposing a global model to manage the bias-variance tradeoff in the context of hedonic house price models* (No. 2022-12). Working Papers in Economics and Statistics.

AKE, I., 2022. *Combining Machine Learning models to predict House Prices* (Doctoral dissertation, Southampton Solent University).

Boyapati, S.V., Karthik, M.S., Subrahmanyam, K. and Reddy, B.R., 2023. An Analysis of House Price Prediction Using Ensemble Learning Algorithms. *Research Reports on Computer Science*, pp.87-96.

Shi, S., 2023. Comparison of Real Estate Price Prediction Based on LSTM and LGBM. *Highlights in Science, Engineering and Technology*, *49*, pp.294-301.

Uyanık, G.K. and Güler, N., 2013. A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, *106*, pp.234-240.

Segal, M.R., 2004. Machine learning benchmarks and random forest regression.

Rathore, S.S. and Kumar, S., 2016. A decision tree regression-based approach for the number of software faults prediction. *ACM SIGSOFT Software Engineering Notes*, *41*(1), pp.1-6.

Song, Y., Liang, J., Lu, J. and Zhao, X., 2017. An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing*, *251*, pp.26-34.

Hancock, J.T. and Khoshgoftaar, T.M., 2020. CatBoost for big data: an interdisciplinary review. *Journal of big data*, *7*(1), pp.1-45.

Avanijaa, J., 2021. Prediction of house price using xgboost regression algorithm. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(2), pp.2151-2155.

Breiman, L., 1996. Stacked regressions. *Machine learning*, *24*, pp.49-64.

Sun, Q. and Pfahringer, B., 2012, December. Bagging ensemble selection for regression. In *Australasian Joint Conference on Artificial Intelligence* (pp. 695-706). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bhandari, H.N., Rimal, B., Pokhrel, N.R., Rimal, R. and Dahal, K.R., 2022. LSTM-SDM: An integrated framework of LSTM implementation for sequential data modeling. *Software Impacts*, *14*, p.100396.

Guillod, T., Papamanolis, P. and Kolar, J.W., 2020. Artificial neural network (ANN) based fast and accurate inductor modeling and design. *IEEE Open Journal of Power Electronics*, *1*, pp.284-299.

HM LAND REGISTRY.2020. UK House Price Index for March 2020 [Online]

Lucian Cook, 2017. Savills Research: "The Global Financial Crisis has fundamentally changed the UK housing market". [Online]

Richardson, J., 2018. *Place and identity: The performance of home*. Routledge.

Gao, G., Bao, Z., Cao, J., Qin, A.K. and Sellis, T., 2022. Location-centered house price prediction: A multi-task learning approach. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(2), pp.1-25.

Lantz, B., 2019. *Machine learning with R: expert techniques for predictive modeling*. Packt publishing ltd.

Vlahovljak, Alen. (2022). House price prediction in the UK.

Lee, J.H., Kim, H.B. and Shim, G.E., 2022. Prediction Model of Real Estate Transaction Price with the LSTM Model based on AI and Bigdata. *International Journal of Advanced Culture Technology*, *10*(1), pp.274-283.

Joe Davies, Kayleigh Mackay, Lili Bui. Understanding towns in England and Wales: town characteristics and house prices, 2019. [Online]