Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 10 (2022) pp. 028-035

A Machine Learning Framework for House Price Estimation

Adebayosoye Awonaike, Seyed Ali Ghorashi[2] and Rawad Hammad[3]

[1] School of Architecture, Computing and Engineering University of East London
London, United Kingdom
U1440387@uel.ac.uk

[2] School of Architecture, Computing and Engineering University of East London
London, United Kingdom
s.a.ghorashi@uel.ac.uk

[3] School of Architecture, Computing and Engineering University of East London

London, United Kingdom
r.hammad@uel.ac.uk

***Abstract*: House prices estimation has been the focus of both commercial and academic researches with various approaches being explored. Depending on the location, size, age, time and other factors, the value of houses may vary. This paper presents a modularized, process oriented, data enabled and machine learning based framework, designed to help the stakeholders within the housing ecosystem to have more realistic estimation of the house prices for decision making. The development of the framework leverages the Design Science Research Methodology (DSRM) and the HM Land Registry Price Paid Data is ingested into the framework as the base transactions data. 1.1 million London based transaction records between January 2011 and December 2020 have been exploited for model design and evaluation. The proposed framework also leverages a range of neighbourhood data including the location of rail stations, supermarkets and bus stops in a cumulative layering approach to explore their possible impact on house prices. Five machine learning algorithms have been exploited and three evaluation metrics have been presented and with a focus on RMSE. Results show that an increase in the variety of parameters enables improved model performance which ultimately will enable decision making. The potential for future work based on this paper can explore the impact of the cumulative introduction of other groups of data on the accuracy of machine learning models designed for the estimation of house prices**

Keywords: House price estimation, Machine learning, Neighbourhood, Random Forest, Price Paid Data.

I. Introduction

Residential housing has been a significant need of human being for a long time. With the average income in the United Kingdom currently around £38,600 for people in full-time jobs and £13,803 for people in part-time jobs [1] the estimated earnings after tax based on the gov.uk tax service calculator is £29,889 and £13,027 respectively. Based on these figures, a 40- year working career for either of these working groups will generate an estimated £1,195,592 and £521,113 respectively of lifetime after-tax income in today's money. Furthermore, with average residential property in the United Kingdom is valued at £231,885 [2] a significant part of our working life is committed to creating the wealth required to own a home. Therefore, though owning a home may probably not be for everyone, everyone needs a place to live and this helps to highlight the range of possible stakeholders in the housing market to include, renters, landlords, investors, developers, housing associations and even local government. To all these stakeholders, being able to estimate the value of residential housing is essential to making critical decisions and ultimately their behaviour as player in the housing market. Hong, Choi and Kim [3] used elapsed year, floor area, floor level of the property and heating system structural factors to determine their impact on price of a house. The outcome showed that elapsed year has a negative correlation with price while area has a positive impact on the price of house. Similarly, Ferlan, Bastic and Psunder [4] observed the same findings on structural factors of a house in Slovenia. An assertion was made that impact of floor level on apartment depends on context i.e., a floor level is a disadvantage if the apartment block has no elevator and but then an incentive if there is an elevator. With housing price being accepted on a larger scale as a research interest as well as a business interest's economic indicator, researchers in [5] proposed a fine-grained model for price predictions. The study describes that how the proposed machine learning model will help to manage existing challenges with property pricing. This fine-grained housing price forecasting model used economic such as GDP, mortgage deposit ratio and social features such as population. For the estimation of house prices in Arlington, North Virginia, a benchmark of Random Forest Machine Learning algorithm with linear regression model was created [6]. It was observed that Random Forest algorithm is able to capture hidden non-linear relations among various features of a house and ultimately give a better house price estimation. Therefore, the resultant model can be used to predict future real estate prices. In the model, the researchers included influencing house prices factors such as zip code, location of the house, year the house was built, house price, and lot size.

A total of 27649 data points were collected from Arlington country, Virginia, USA in 2015. All the data were of a single-family house. Random forest algorithm performed better in terms of R-square and RMSE. While there are quite a few academic and commercial machine learning based research exist on the subject of house price estimation, there has not been enough focus on the design of a robust framework that continues to learn through batch data ingestion.

In this paper we proposed a framework by focusing on the design of a modularized, process-oriented, data-enabled, machine learning-powered framework. This framework exploits publicly available data and enriches the HM Land Registry's price paid data by geocoding it (making it spatially enabled) and blending it with neighbourhood features such as distance from nearest rail stations, bus stops and supermarkets. Leveraging the capabilities of five different machine learning algorithms; LightGBM for accuracy, efficiency and low computational cost [7], Random Forest for strong performance, ability to handle categorical data with multiple levels, and working adequately with missing data [6], XGBoost for scalability [8] and Hybrid Regression and Stacked Generalization being ensemble of algorithms [9], [10] and [11]. This paper further takes a cumulative multi-feature layering approach to explore the impact of groups of parameters shown in figure 1 on the accuracy of these machine learning algorithms.

The rest of this paper is organized as follows. Section II focuses on the methodology deployed including data collection In Section III modelling and results are discussed while section IV explains model optimization and evaluation. Finally, we conclude in section V.

II. Methodology

A variety of research methods that have been explored in existing researches including experimental [12], comparative study [13] and systematic sampling [14]. Since this paper aims to explore a feature layering approach for the design of a multi-feature house prices estimation framework, the Design Science Research Methodology is explored [15]. As shown in figure 1, its process includes (i) problem identification and motivation, (ii) definition of the objectives for a solution, (iii) design and development, (iv) demonstration, (v) evaluation, and (vi) communication. The first three will be discussed in this section while remaining three will be spread across section III to V.

Figure 1: Framework Design Methodology

A. Problem Definition

A number of machine learning driven and multi-data-enabled frameworks already exit for the estimation of house prices however, there is a lack of research to understand the potential impact cumulative feature layering on the selection of machine learning algorithms and the performance of models.

B. Definition of Objectives

The objectives of this paper include

a) Develop a modularized framework to ensure ease of development

b) Use an ensemble of machine learning techniques and identify which suites best to estimate house prices based on available data

c) Explain the impact of data variety on machine learning models

C. Design and Development

The design and development explore data collection, the creation of the modules and pipeline to ensure ease of development and reusability.

1) Data Collection

All the datasets exploited in this paper are publicly available. These are (i) price paid data, (ii) Office of National Statistics – National Statistics Postcode Lookup (ons nspl) product, (iii) supermarkets, (iv) bus stops and (v) rail stations. [22] These make up the first and second tiers of data exploited as shown in figure 2.2.

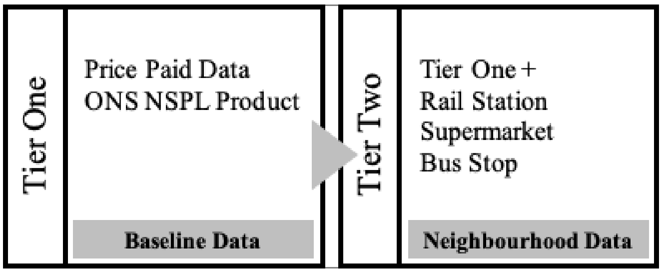

Figure 2: Feature Tiers for a cumulative multi-feature data-enabled framework

Tier 1

The HM Land Registry Prices Paid Data is described as "the official house price dataset in England" [16]. published by HM Land Registry. The single download file for this dataset contains over 25 million records with 16 variables representing information on all property sales transactions in England and Wales from 1 January 1995 to date [17]. In this paper we have exploited 1.1 million records which represents all the London based transactions. The ONS produce two main postcode products. These are (i) ONS Postcode Directory (ONSPD) and (ii) National Statistics Postcode Lookup (NSPL). These products are widely used by a range of customers including central and local government, commercial organizations and academia [18]. The ONSPL has been used in this paper to create a geocoded version of the HM Land Registry price paid data.

Tier 2

The "GB Rail Stations dataset" is a list of all the train stations in the United Kingdom. It is made up of 2,569 records and 39 variables including the volume of passengers travelling through each station either as the start or end of their journeys or even as an interchange is captured. This paper takes a scientific approach to explore the possible impact of the distance to the nearest train stations on house prices.

Supermarkets are also opening up in multiple location and in different formats, thereby providing customers with choice. With so many new stores, it becomes relatively easier to know where the competition is and consequentially, the new markets being targeted by retailers. Therefore, this paper also explores the impact of the distance to supermarkets on house prices using United Kingdom Supermarket "Retail Points", a dataset of supermarkets. This Geolytix data has 16,991 records and 17 variables.

The bus stop dataset is published to the National Public Transport Data Repository by the Department for Transport. The data in this repository is available from October 2004 to October 2011. However, it is now static and superseded by the Traveline National Dataset.

Table 1: Publicly available datasets collected and exploited in machine learning models

Tier	Data	Records	Source
1	Price Paid Data	1,096,000	gov.uk
	ONS NSPL	2,661,131	ons.gov.uk
2	Rail Station	2,569	doogal.co.uk
	Supermarket	16,991	geolytix.com
	Bus Stop	406,873	data.gov.uk

2) Modular Programming

To introduce a robust solution that ensures ongoing and future development is done more quickly, a modular approach has been followed in this paper. A module approach has been followed to develop this framework. This module approach aims at applying software engineering principles including low coupling and high cohesiveness to produce a maintainable software tool that is scalable in the future.

Figure 3 shows the modules that define this framework and they include (i) assets – holding all raw data to be ingested as detailed in figure 1, (ii) data ingestion – comprising of functions designed to ingest all research data from the asset module and also extract the specific records required as baseline, (iii) data processing – this is made up of the functions designed for data cleansing and initial exploratory data analysis, (iv) features engineering – this caters for the engineering of features in datasets across both tiers

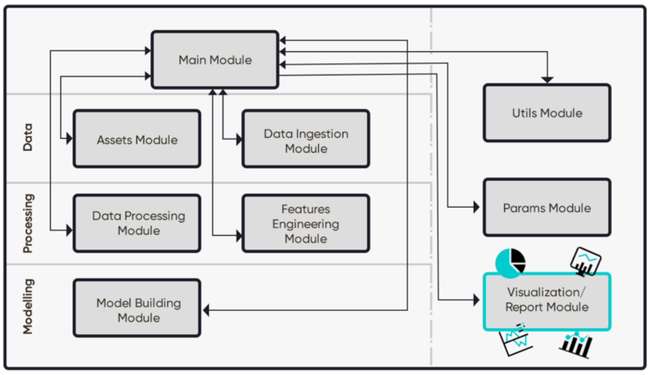

Figure 3: Modules of the Multi-feature House Price Estimation Framework

(v) model building – this is where the machine learning algorithms are exploited on the training data, test data and validation data (vi) params – this module serves as the store for multiple dictionaries created to hold groups of data that belong to the same tier, (vii) utils – this comprises of classes and functions designed to handle processing tasks like the label encoding of categorical features and feature union of numerical features, and (viii) Main is the primary environment where the functions and classes across other modules are called into action.

III. Modelling and Results

Baseline models were created using five modelling techniques but using default parameters (i.e., no tuning at this point). These are (i) Light GBM, (ii) Random Forest, (iii) XGBoost, (iv) Hybrid Regression and (v) Stacked Generalization. These regression algorithms have been selected because of the speed of learning, the handling of overfitting to improve accuracy and high flexibility.

Light GBM (Light Gradient Boosting Machine) is a machine learning algorithm based on decision trees. It has wide application in real-life such as ranking, classification and tasks based on machine learning. Its development focuses on performance and scalability. It's advantages include sparse optimization, early stopping, parallel training, bagging, regularization as well as multiple loss functions. Exclusive Feature Bundling (EFB) and Gradient-based one-side sampling (GOSS) are the two powerful techniques used by LightGBM to improve accuracy, efficiency and memory consumption as well as speed [7].

Random Forest is a machine learning (ML) algorithm based on weak learner decision trees [6]. The ensembling manner of the decision trees eliminates instability problems as well as overcome the high variance of decision trees. Since decision trees are generated by random sampling method, hence the name is random forest.

XGBoost exists as an open-source package for tree boosting and can be described as a machine learning system that is scalable. Its scalability feature allows the users to quickly define their objectives. This algorithm is also very fast as it performs parallel computation, it accepts a wide array of inputs, it has sparsity, customization and acceptable performance [8].

Hybrid Regression is and ad-hoc and user-defined ML method. In this regard, hybrid regression entails combining two or more ML methods to develop a unique ML method. The combined outcome of these ML methods is far better than the results of each ML method by itself. For instance, [9] tested houses Prices in Beijing with a model consisting of 33.33% Random Forest, 33.3% LightGBM and 33.3% XGBoost. The hybrid model achieved a better result of RMSE of 0.14969 far better that each algorithm run solely. Similarly, [10] realized a better overall result of a hybrid model consisting 65% Lasso and 35% Gradient Boosting.

Stacked Generalization was introduced by Wolpert [11] and it's a Python based package. The main idea behind this method is to use predictions of the previous models as features for the present model. Stacked Generalization uses K-fold cross-validation to avoid overfitting. For instance, [9] used 2-level stacking architecture, the first stack comprised on Random Forest and LightGBM while the second stack comprised on XGBoost to predict the house prices. They also noted that the combined results are not as impressive as the Hybrid Regression

Table 2 show the results for all models using the default parameters for each algorithm on only the geocoded HM Land

Registry Price Paid Data while table 3 show the results for all models using the default parameters for each algorithm after a layer of neighborhood data (shown in figure 2) have been introduced. In this paper, the introduction of the tier 2 datasets is described as cumulative multi-feature layering. Figure 4 to 8 show the plot of predicted versus actual house prices from models using the default values of the five machine learning algorithms listed above.

Table 2: Tier 1 modelling results using default parameters (RMSE)

Model	Train	Test
Light GBM	2399452.94	3520071.29
Random Forest	1417480.98	3702246.49
XGBoost	2224870.41	3705299.63
Hybrid Regression	1943280.22	3552304.53
Stacked Generalisation	2605422.57	3705613.80

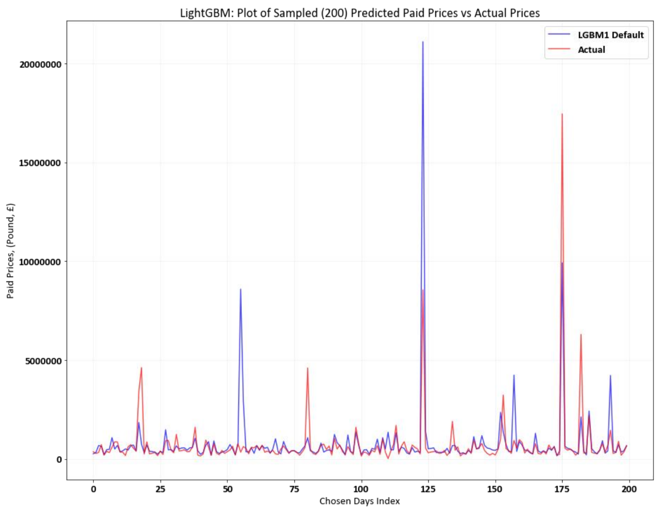

Figure 4: Plot of LightGBM default model output vs actual – Tier 1 (first 200 houses)

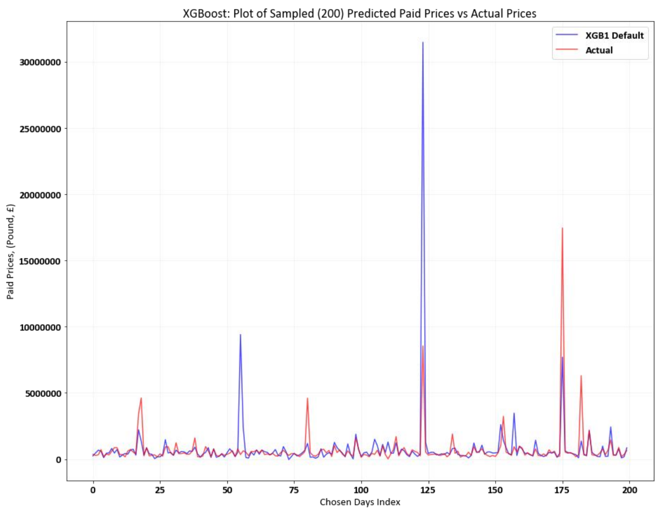

Figure 5: Plot of XGBoost default model output vs actual – Tier 1 (first 200 houses)

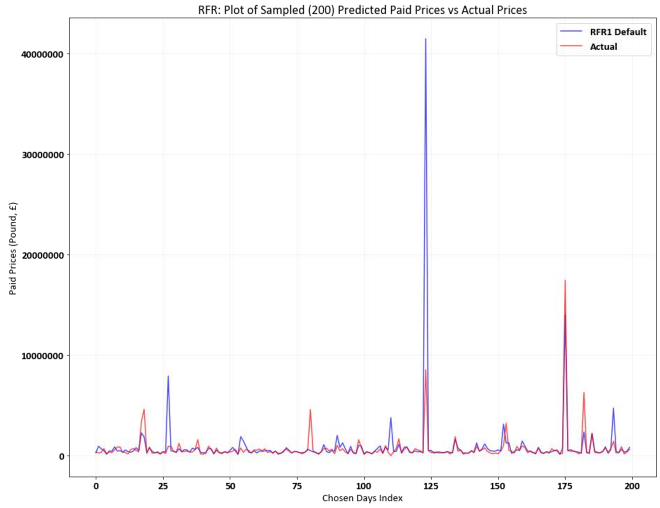

Figure 6: Plot of Random Forest default model output vs actual – Tier 1 (first 200 houses)

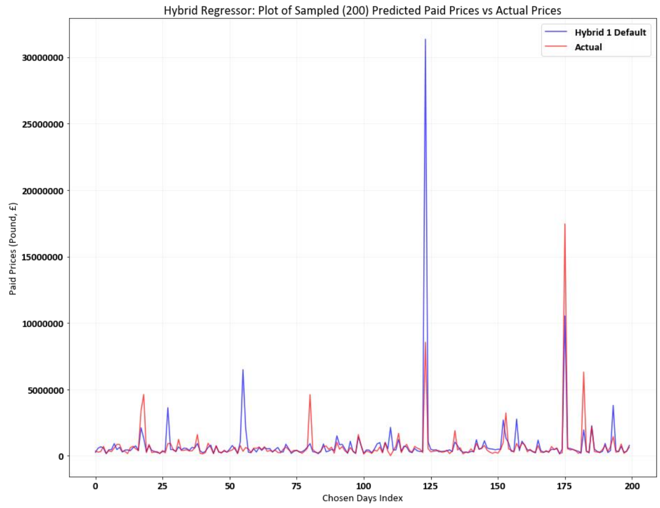

Figure 7: Plot of Hybrid Regression default model output vs actual – Tier 1 (first 200 houses)

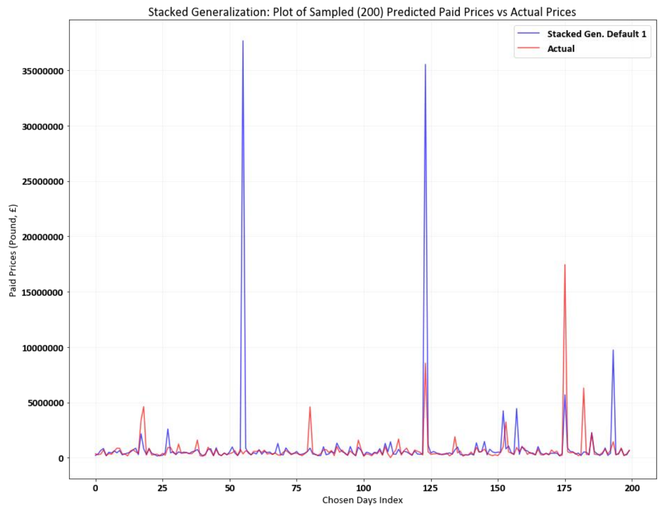

Figure 8: Plot of Stacked Generalisation default model output vs actual – Tier 1 (first 200 houses)

RMSE, Root Mean Square Error is a common metric used to measure the error of a model predicting quantitative data [19]. It estimates the standard deviation of an observed value from the model prediction. According to [19], the observed value is equal to the sum of the predicted value and predictably distributed random noise with mean zero. If the noise is negligible as estimated by RMSE, the model is assessed as good at predicting the observed data. However, if RMSE is large, it means that the model is not accounting for important features in the data.

Table 3: Tier 2 modelling results using default parameters (RMSE)

Model	Train	Test
Light GBM	2248053.95	3477427.88
Random Forest	945781.32	3492294.16
XGBoost	2224870.41	3705299.63
Hybrid Regression	1640064.48	3416892.70
Stacked Generalisation	2345842.61	3631184.76

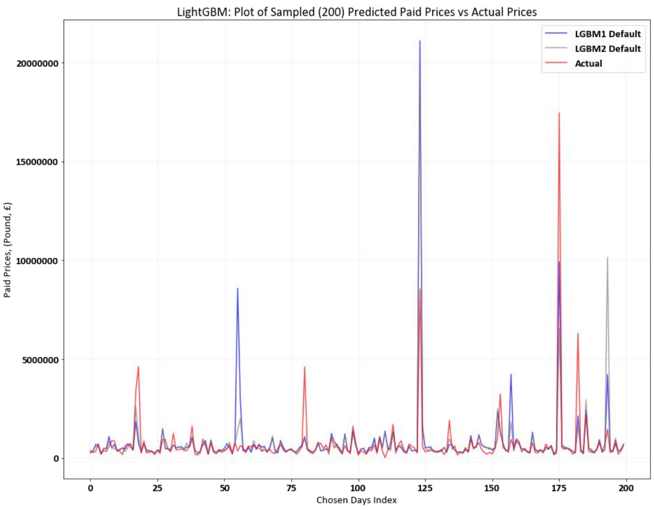

Figure 9: Plot of LightGBM default model output vs actual – Tier 2 (first 200 houses)

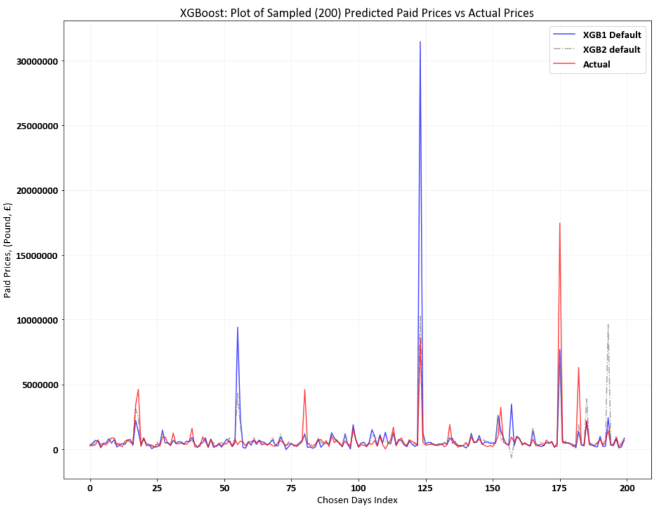

Figure 10: Plot of XGBoost default model output vs actual – Tier 2 (first 200 houses)

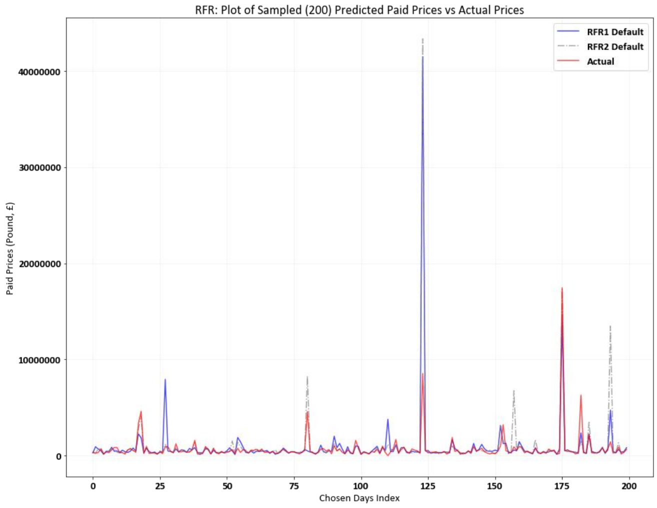

Figure 11: Plot of Random Forest default model output vs actual – Tier 2 (first 200 houses)

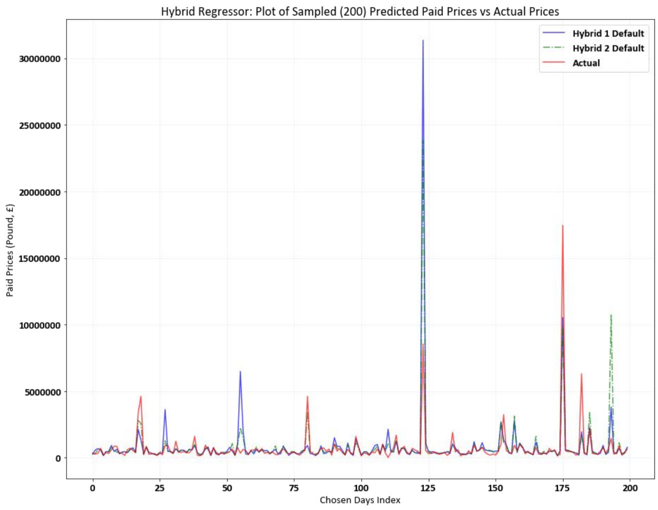

Figure 12: Plot of Hybrid Regression default model output vs actual – Tier 2 (first 200 houses)

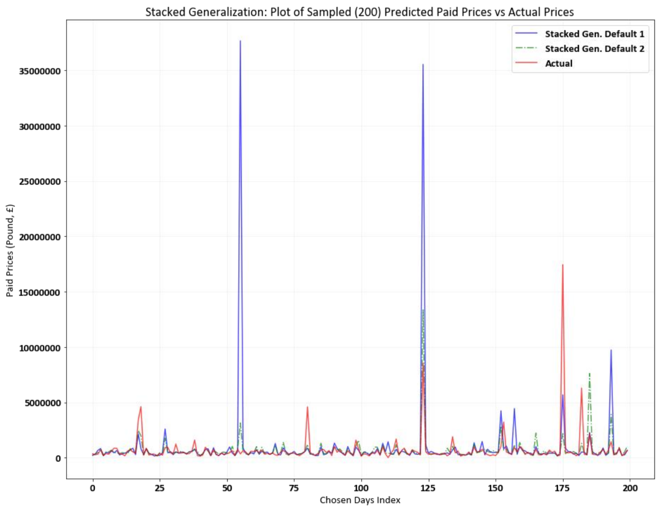

Figure 13: Plot of Stacked Generalisation default model output vs actual – Tier 2 (first 200 houses)

As shown in table 4, a reduction in RMSE values for 80% of the models in Tier 2 compared with Tier 1 shows an improvement in model performance based on the cumulative introduction of neighborhood features. Random Forest model showed the most improvement followed by Hybrid regression, stacked generalisation, and LightGBM respectively while XGBoost had no change. This also shows that machine learning algorithms respond the features in different ways.

Table 4: A comparison of modelling results for Tier 1 and Tier 2 with a focus on RMSE

Model	Tier 1	Tire 2	%
Light GBM	3520071.29	3477427.88	1.2%
Random Forest	3702246.49	3492294.16	5.7%
XGBoost	3705299.63	3705299.63	0.0%
Hybrid Regression	3552304.53	3416892.70	3.8%
Stacked Generalisation	3705613.80	3631184.76	2.0%

IV. Model Optimisation and Evaluation

Model optimization in machine learning is, one of the most challenging aspects of the implementation of ML solutions. There is immense attention given to deep learning theories and machine learning to achieve the optimization of models. There exist two types of parameters in models of machine learning; model parameters - which possess the ability to be initiated and consequently updated through data learning and hyper-parameters - which have to be set before the training of the machine learning model since they are associated with the configuration of the machine learning model [20].
Hyperparameters were calculated using the 'Bayesian Optimization' method and applied to three of the baseline models i.e. (i) Light GBM, (ii) Random Forest and (iii) XGBoost.

Table 5: Hyperparameters used based on Bayesian Optimisation method

Model	Tier 1	Tier 2
Light GBM	{'colsample_bytree': 0.7139109116423488, 'learning_rate': 0.1133802923850488, 'max_depth': 1, 'min_child_samples': 300.0, 'n_estimators': 2, 'num_leaves': 400.0, 'reg_alpha': 2.697524711103592, 'reg_lambda': 7.792201841119942, 'subsample': 0.850664221897626, 'subsample_for_bin': 26000.0}	{'colsample_bytree': 0.9298462716109409, 'learning_rate': 0.10075805121869513, 'max_depth': 0, 'min_child_samples': 300.0, 'n_estimators': 3, 'num_leaves': 600.0, 'reg_alpha': 2.671148237480234, 'reg_lambda': 0.31242913248603976, 'subsample': 0.8930209167711038, 'subsample_for_bin': 30000.0}
Random Forest	{'max_depth': 16.80565381273467, 'min_samples_leaf': 4.548341736478191, 'n_estimators': 302.9997495928546}	{'max_depth': 19.633732543557223, 'min_samples_leaf': 4.900530193117771, 'n_estimators': 366.02562034116534}
XGBoost	{'colsample_bytree': 0.62, 'gamma': 0.32, 'learning_rate': 0.02, 'max_depth': 2, 'min_child_weight': 10.0, 'n_estimators': 4, 'subsample': 0.54}	{'colsample_bytree': 0.73, 'gamma': 0.31, 'learning_rate': 0.09, 'max_depth': 2, 'min_child_weight': 8.0, 'n_estimators': 2, 'subsample': 0.72}

The hyperparameters used are shown in table 5 according to [21], there are four common methods explored for the hyperparameter optimization of machine leaning models and Bayesian model- based Optimization is assessed as most efficient. The others include random search, grid search and manual.
Table 6 and Table 7 show the model results for three optimized models using three different metrics. Using RMSE, table 8 shows a consistent improvement for Random Forest and LightGBM models in Tier 2 compared to Tier 1. This also confirms that the introduction of neighborhood features leads to improved performance of the Random Forest model. Figure 14 and 15 show the results Tier 1 features using optimised parameters for Light GBM and Random Forest models while figure 16 and 17 show the results Tier 2 features using optimised parameters for Light GBM and Random Forest models.

Table 6: Modelling results using optimised parameters (Tier 1) - RMSE

Model	Train	Test
Light GBM	2706799.10	3668646.78
Random Forest	2222173.82	3523322.99
XGBoost	2597208.98	3563082.33

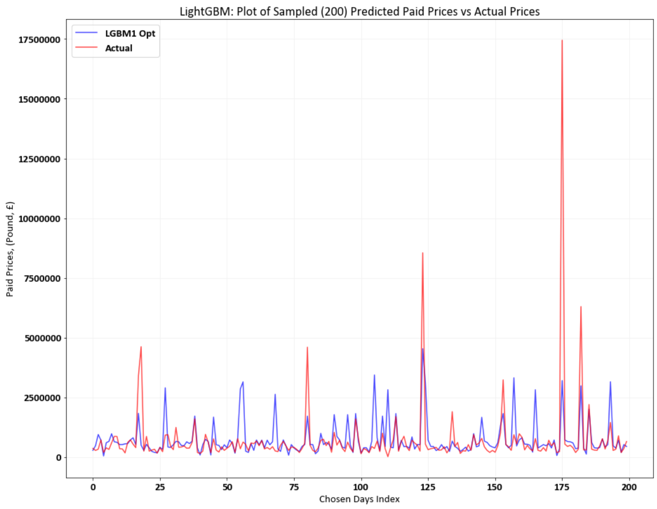

Figure 14: Plot of LightGBM optimised model output vs actual – Tier 1 (first 200 houses)

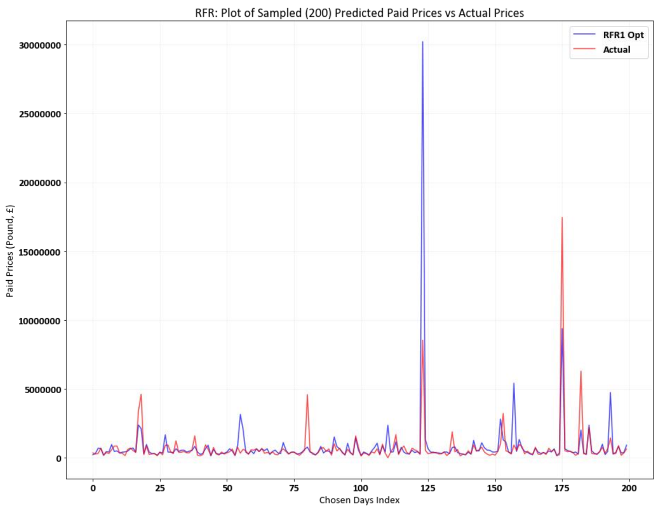

Figure 15: Plot of Random Forest optimised model output vs actual – Tier 1 (first 200 houses)

Table 7: Modelling results using optimized parameters (Tier 2) - RMSE

Model	Train	Test
Light GBM	2639464.50	3611587.57
Random Forest	1980820.53	3415804.71
XGBoost	2762658.11	3714580.93

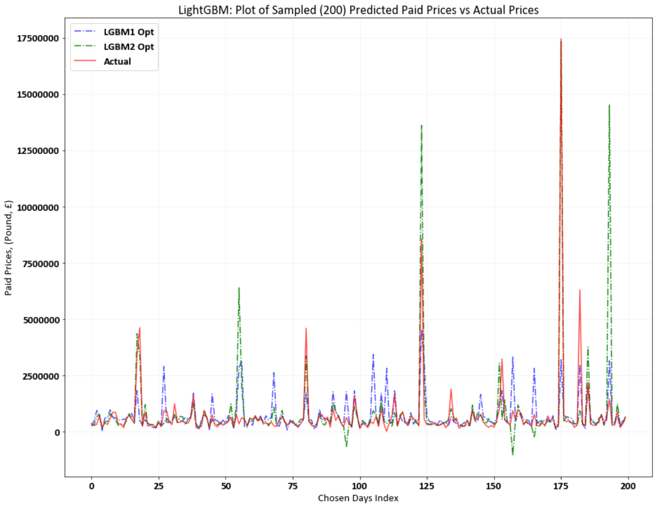

Figure 16: Plot of LightGBM optimised model output vs actual – Tier 2 (first 200 houses)

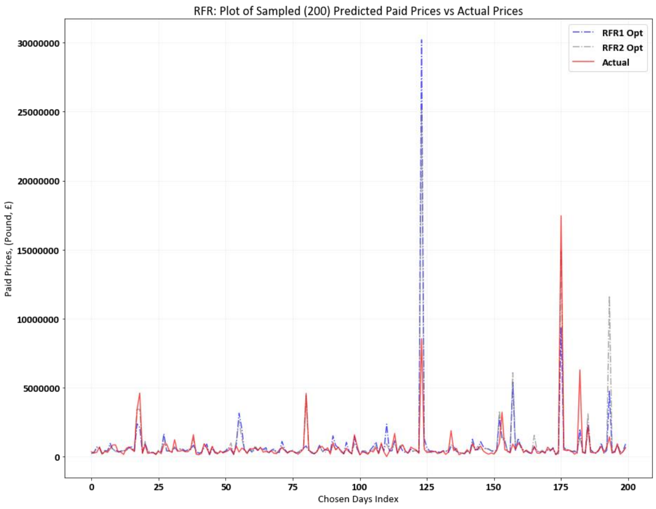

Figure 17: Plot of Random Forest optimised model output vs actual – Tier 2 (first 200 houses)

Table 8: A comparison of modelling optimization results for Tier 1 and Tier 2 with a focus on RMSE

Model	Tier 1	Tier 2
Light GBM	3668646.78	3611587.57
Random Forest	3523322.99	3415804.71
XGBoost	3563082.33	3714580.93

V. Discussion and Conclusion

This paper has presented a modularized, process-based, data driven and machine learning enabled framework for housing price estimation. It is comprised of sixteen modules leveraging the design science research methodology and five publicly available datasets grouped into two tiers. Five Machine Learning methods and techniques including LightGBM, Random Forest, XGBoost, Hybrid Regression and Stacked Generalization have been exploited and analysed for optimal estimations.

Although these methods and techniques produced some intriguing results, they all have their advantages and disadvantages. Though Random Forest model results using RMSE as an evaluation metric shows the best performance improvement, the performance of models for Hybrid Regression, Stacked Generalisation and LightGBM also improved as a result of the cumulative introduction of tier 2 features. These results also reveal the behaviour of machine learning models in response to changing variety of models. However, it will be interesting to see how these models and the evaluation metrics respond to an increase in the variety of features within the framework.

[22] With continuous improvement of the estimated house price, developers or builders and other stakeholders withing the housing ecosystem will have access to data-enabled insights on where to build, what kind of homes to build, where to buy or invest and where to rent owing to the unique features that define different geographic areas. [22] Investors and landlords will also have better insights to calculate their return on investment and new or serial tenants are able to identify geographic areas that best suit their personal preferences. Further work will be required to explore the potential impact of ingesting an additional layer of features.

References

[1] Office of National Statistics. "Employee Earnings in the UK: 2020." ONS. (accessed 30 May 2021, 2021).

[2] HM Land Registry. "UK House Price Index for March 2020." HM Land Registry. (accessed 5 January 2021, 2021).

[3] J. Hong, H. Choi, and W.-S. Kim, "A House Price Valuation Based on The Random Forest Approach: The Mass Appraisal of Residential Property in South Korea," *International Journal of Strategic Property Management*, vol. 24, no. 3, 2020.

[4] N. Ferlan, M. Bastic, and I. Psunder, "Influential Factors on the Market Value of Residential Properties," *Engineering Economics*, vol. 28, no. 2, p. 10, 2017.

[5] C. Ge, Y. Wang, X. Xie, H. Liu, and Z. Zhou, "An Integrated Model for Urban Subregion House Price Forecasting: A Multi-Source Data Perspective," presented at the In 2019 IEEE International Conference on Data Mining (ICDM), 2019.

[6] C. Wang and H. Wu, "A New Nachine Learning Approach to House Price Estimation," *New Trends in Mathematical Sciences*, vol. 6, no. 4, pp. 165-171, 2018.

[7] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *Advances in neural information processing systems*, vol. 30, pp. 3146-3154, 2017.

[8] T. Chen and C. Guestrin, "Xgboost: A Scalable Tree Boosting System," presented at the International Conference on Knowledge Discovery and Data Mining, 2016.

[9] Q. Truong, M. Nguyen, H. Dang, and B. Mei, "Housing Price Prediction via Improved Machine

Learning Techniques," *Procedia Computer Science*, vol. 174, pp. 433-442, 2020.

[10] S. Lu, Z. Li, Z. Qin, X. Yang, and R. Goh, "A hybrid Regression Technique for House Prices Prediction," in *International Conference on Industrial Engineering and Engineering Management*, 2017, pp. 319-323.

[11] D. H. Wolpert, "Stacked Generalization.," *Neural Networks: The Official Journal of the International Neural Network Society*, vol. 5, no. 2, pp. 241-259, 1992.

[12] B. Park and J. K. Bae, "Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data," *Expert Systems with Application*, vol. 42, no. 6, 2014.

[13] C. R. Madhuri, G. Anuradha, and M. V. Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," presented at the International Conference on Smart Structures and Systems, Chennai, India, 2019.

[14] J. R. Rico-Juan and P. Taltavull de La Paz, "Machine Learning with Explainability or Epatial Hedonics Tools? An Analysis of the Asking Prices in the Housing Market in Alicante, Spain," *Expert Systems with Applications*, 2021.

[15] R. Hammad, "A Hybrid E-Learning Framework: Process-Based, Semantically-Enriched and Service-Oriented," Doctor of Philisophy, Faculty of Environment and Technology, University of West England, 2018.

[16] B. Chi, A. Dennett, T. Oléron Evans, and R. Morphet, "Creating a New Dataset to Analyse House Prices in England," *CASA Working Paper*, vol. 213, 2019.

[17] HM Land Registry. "Statistical Dataset Price Paid Data." HM Land Registry. https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads (accessed 30/05/2021.

[18] Office of National Statistics. *National Statistics Postcode Lookup*. [Online]. Available: www.ons.gov.uk

[19] J. Moody. "What does RSME really mean?" https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e (accessed 11/09/2021, 2021).

[20] A. Zhang. "Evaluating Machine Learning Models." https://www.oreilly.com/library/view/evaluating-machine-learning/9781492048756/ch04.html (accessed 28/06/2021.

[21] W. Koehrsen. "A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning." Towards Data Science https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f (accessed 06/06/2021, 2021).

[22] A. Awonaike, S.A. Ghorashi, R. Hammaad, (2022). A Machine Learning Framework for House Price Estimation. In: Abraham, A., Gandhi, N., Hanne, T., Hong, TP., Nogueira Rios, T., Ding, W. (eds) Intelligent Systems Design and Applications. ISDA 2021. Lecture Notes in Networks and Systems, vol 418. Springer, Cham. https://doi.org/10.1007/978-3-030-96308-8_90

Author Biographies

Adebayosoye Awonaike is a PhD in Data Science candidate at the University of East London, United Kingdom. He has a BSc in Geology at the Olabisi Onabanjo University, Ogun State, Nigeria in 2003, and MSc in Geographical Information Systems with Remote Sensing at University of Greenwich, London United Kingdom in 2007.

Seyed Ali Ghorashi was born at Tehran, Iran in 1969. He received his BSc and MSc degrees in Electrical Engineering from the University of Tehran, Iran in 1992 & 1995 respectively, and his PhD degree in Communication Networks from King's College London, UK in 2004. He has worked for Samsung Electronics (UK) Ltd, Shahid Beheshti University, Middlesex University, and University of East London. He is a senior member of IEEE, holds US and international patents and has published over 130 technical papers mainly related to the applications of optimization, artificial intelligence and machine learning in positioning, internet of things and wireless communications.

Rawad Hammad is a Programme Leader for MSc Computing, Programme Leader for MSc Digital Education, Technology Enhanced Learning Research Group Leader, and a Senior Lecturer in Computer Science and Digital Technologies at the University of East London. Rawad has extensive experience on Software Engineering, Technology Enhanced Learning (TEL), Artificial Intelligence in Education, and Smart Health research and practice. Rawad is an executive committee member of the International Society of the Artificial Intelligence in Education (AIED) and committee member of different research bodies.