KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

COLLEGE OF SCIENCE
DEPARTMENT OF MATHEMATICS

APPLICATION OF MACHINE LEARNING IN PREDICTING HOSTEL PRICES: A Case Study of KNUST

BY

ANNAN JESSE
BOATENG GLORIA MAAME SERWAA
DENTE-QUARSHIE OBENG KWAME

ABSTRACT

Machine learning, which dates back to the 1950s, aims to make accurate predictions with unseen (similar) data based on patterns discovered in existing data. In the real estate market, machine learning algorithms have been deployed in either determining house price indexes or estimating house sale prices. The latter is being studied the most by considering factors such as location, population, proximity to a nearby station, zip code, and many more. However, research on hostel price prediction is rare, if not nonexistent. To help bridge the gap, we will explore the impact of hostel features using three machine learning algorithms: multiple linear regression, ridge regression, and neural network, in predicting hostel prices. Empirical results support the potential of machine learning algorithms on the hostel market, with all R^2 greater than 0.75.

ACKNOWLEDGMENTS

The completion of this project could not have been possible without the assistance of numerous people. Therefore, we would like to express our deep appreciation and indebtedness to our supervisor, Dr. Charles Sebil, whose insightful leadership and knowledge helped us complete this project successfully.

We would also like to express our deep appreciation to Dr. Emmanuel Kofi Gavu and Ms. Deborah Dormah Kanubala for their valuable suggestions and ideas towards the completion of this project and to all the lecturers we have encountered. We thank you for allowing us to glean knowledge from you throughout our stay here.

Finally, we would like to thank all our relatives, family, and friends who supported us in one way or another. Above all, we would like to thank God Almighty for always showering us with his blessings.

Contents

List of Abbreviations

ML Machine Learning

MLR Multiple Linear Regression

RR Ridge Regression

NN Neural Network

ReLU Rectified Linear Unit

MAR Missing At Random

MCAR Missing Completely At Random

PRSS Penalized Residual Sum of Squares

MAD Median Absolute Deviations

MAE Mean Absolute Error

RMSE Root Mean Squared Error

CoS College of Science

PHA Private Hostels Association

SRC Students' Representative Council

KNUST Kwame Nkrumah University of Science and Technology

Chapter 1

INTRODUCTION

1.1 Background

Renting a hostel is undoubtedly one of the most important decisions a student makes during their entire stay on campus. The price of these hostels depends on a wide variety of factors, ranging from location to the number of beds in a room, access to the shuttle, and many more. As the population of students increases, traditional hostel price predictions based on hostel price comparisons lacking an accepted standard can no longer be employed to estimate hostel prices. Therefore, the availability of a hostel price prediction model helps fill an important information gap (Calhoun, 2003) and boosts the hostel market efficiency. Machine Learning (ML) is one of the cutting-edge techniques that can be utilized to identify, interpret, and analyze hugely complicated data structures and patterns (Ngiam and Khor, 2019).

1.1.1 Machine Learning Overview

Machine Learning (Douglass, 2020) is the science of programming computers to learn from data. The three major types of ML are supervised, unsupervised, and reinforcement learning. In supervised learning, the ML is guided with desired inputs and outputs by a human operator; the algorithm learns from the data and makes predictions. A typical supervised learning task is regression. Unsupervised learning algorithms find insight in unlabeled training data and organize the data in some way to describe its structure; a typical task is clustering. Reinforcement learning (Shweta Bhatt, 2018) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its actions and experiences; the commonly used algorithm is Q-learning.

1.2 Problem Statement

In developed countries, ML has been implemented successfully to estimate real estate prices. However, in Ghana, the application of ML is rare and few and far between on the topic of hostels. The lack of adequate data (Owusu-Ansah, 2012) has made it difficult, if not impossible, to develop efficient or systematic hostel pricing policies through modeling the hostel market. As a result, hostel managers always overprice their hostel rooms. This study aims to derive valuable insight into KNUST's hostel market through analyzing a real historical dataset. It seeks useful models to illustrate how ML algorithms can be utilized to predict hostel prices given a set of its features.

1.3 Significance

Our models, if accepted, could allow students or hostel managers to make better decisions. In addition, it could benefit the projection of future hostel prices and the policy-making process for the hostel market.

1.4 Limitations

- This paper proves the competence of ML algorithms in the hostel market. Although it aims to assist policymakers in projecting future hostel prices, it does not, however, explicitly forecast hostel prices.
- 70 out of 105 recommended hostels by Kwame Nkrumah University of Science and Technology (KNUST); were considered for this study.
- Although several students rent homes (homestels), this study only focused solely on the prices of private hostels around KNUST.

1.5 Outline

The next parts of this paper are constructed as follows: In Chapter 2, existing literature on housing market prediction applying different ML algorithms will be reviewed. Chapter 3 explores the dataset, explains how to transform it into cleaned data, and introduces the various supervised ML algorithms implemented. Chapter 4 presents our empirical results and the conclusion deduced in Chapter 5.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

Previous studies on the real estate market using ML approaches can be categorized into two groups: the trend forecasting of house price indexes and house price valuations (Phan, 2019). Since house prices are strongly correlated to other factors such as location, area, and population (Kamal et al., 2021), it requires other information apart from the house price index to predict individual house prices (Truong et al., 2020).

2.2 Related Work

(Gavu and Owusu-Ansah, 2019), one of the very few papers to model the residential rental housing market in Ghana, implemented the hedonic price model to estimate house prices in the Accra Metropolis, Adenta, Ga East, La Dade Kotopon, and La Nkwantanang Madina areas while exploring the existence of submarkets.

(Pan and Zhong, 2019) implemented and compared the performance of Ridge, Lasso, Multiple Linear Regression, Neural Nets, and Random Forest. In their findings, neural networks proved to be the most accurate in estimating house value.

(Selim, 2009) and (Limsombunchai, 2004) papers offer a comparative approach between hedonic model and artificial neural network, although, on different datasets, results show that artificial neural network performs significantly better. On the other hand, (Nguyen and Cripps, 2001) compared multiple linear regression and artificial neural networks. Empirical results show that the performance of artificial neural networks improved as the data size increased.

L_1 and L_2 regularization was implemented by (Xin and Khalid, 2018) on a housing dataset from 2006 to 2010, lasso regression (L_1) produced much better predictions; while (Madhuri et al., 2019), involved Multiple Linear, Elastic Net, Gradient Boosting, and Ada Boost regression together with L_1 and L_2 regularization for estimating house price value; with a score of approximately 0.92, Gradient Boosting proved superior.

(Thamarai and Malarvizhi, 2020) applied Multiple Linear Regression and Decision Tree Regression to a housing dataset. Comparatively, the performance of multiple linear regression produced better results.

(Vineeth et al., 2018) modeled a Kaggle housing dataset utilizing 19 regression algorithms to help consumers find a price for their soon-to-be house without consulting a real estate agent (broker). Cat boost was the best performing model based on RMSE, with a value of $2.604e + 04$.

Chapter 3

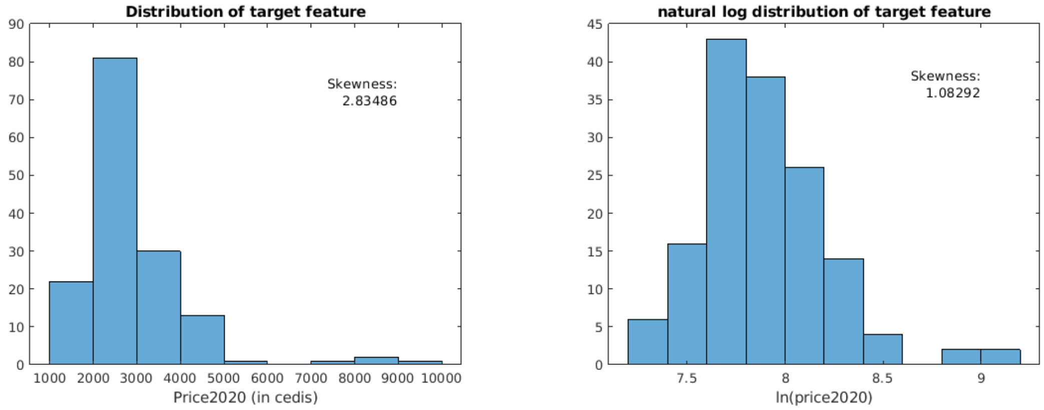
METHODOLOGY

3.1 Introduction

In this chapter, we explore our data, perform data preparations, and feature engineering on the dataset, then we build, compare, and evaluate three models: neural network, linear, and ridge regression.

3.2 Exploratory Data Analysis

Exploratory data analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypotheses, and check assumptions with the help of summary statistics and graphical representations (Patil, 2018). Visualizations not shown in this section are allocated in the appendix. Figure 3.1a shows skewed data of the target feature and 3.1b shows the transformation of the target feature using a natural log, which is deemed to have a normal distribution.



(a) Distribution of price2020

(b) Distribution of ln(price2020)

Figure 3.1: Histograms of target feature 1

3.3 Data Preparation

Data preparation is the act of manipulating raw data (which may come from disparate data sources) into a form that can readily and accurately be analyzed (Friedland). In light of that, we first deal with missing data and remove inputs that behave like outliers. Next, we select important features and feature engineer the categorical features.

3.3.1 Missing Data

Most ML algorithms cannot work with missing features (Douglass, 2020). Missing data are unobserved values that would be meaningful for analysis if observed. In other words, a missing value hides a meaningful value (Little and Rubin, 1988).

Table 3.1: Missing Data in hostels dataset

Feature	grade	rank	price2018	price2019
%Missing	3%	3%	89%	64%
numMissing	5	5	135	97
numTotal	242			

The two types of missingness observed in our dataset are Missing At Random (MAR) and Missing Completely At Random (MCAR). The grade is of type MCAR since respondents chose to skip answering; hence they were fixed using the imputation method. Rank is of type MAR considering its value depends on the grade. price2018 and price2019 are of type MCAR because, in most cases, respondents were not present in their hostels during those academic years and thus discarded completely.

3.3.2 Outliers

By default, an outlier is a value that is more than three scaled Median Absolute Deviations (MAD) away from the dataset (Mathworks). Outliers can easily affect the performance of the model(Parashar, 2021). Hence, we use 99% of our dataset to reduce their effect on the data (and models).

3.3.3 Feature Selection

Feature selection is a subfield within ML aimed at creating accurate models by excluding irrelevant features and including only relevant features (Jaiantilal, 2013). Features are selected based on their scores in various statistical tests (e.g., Pearson's correlation) for their correlation with the outcome variable.

3.3.4 Feature Engineering

Every ML model is based on some mathematical concept, so we encode every categorical value into a numerical value. Features without inherent order, such as study_room were one-hot-encoded, and features with inherent order, such as rank, were label encoded. Also, beds and post_code features were deemed categorical since they contain 4 and 21 distinct responses, respectively, with no inherent order. Therefore, we one-hot-encoded the two features.

3.3.5 Data Splitting

Data splitting is the process of partitioning our available data into a train set (for training all models) and a test set (to evaluate the models). Using the cvpartition object together with the Mersenne Twister random number generator (i.e., rng (1)) in Matlab (R2021a), the cleaned dataset will be randomly partitioned, with 65% as a training set and 35% as a test set.

3.4 Model Selection

3.4.1 Criteria to Measure Performance

The test set were evaluated using: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2)

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.1}$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}SSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.2}$$

Coefficient of Determination (R^2)

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad SSE = \sum_{i=1}^{n}(y_i - \hat{y})^2$$

$$R^2 = 1 - \frac{SSE}{SST} \tag{3.3}$$

3.4.2 Predictive Modeling

Linear Regression

Linear regression (Chatterjee and Simonoff, 2013) is a linear approach to modeling the relationship between a response and one or more explanatory variables. Given a dataset $\{x_{1i}, x_{2i}, ..., x_{pi}, y_i\}$ of n sets of observations, it is assumed that these observations satisfy a linear relationship,

$$y_i = \beta_o + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \epsilon_i = X\beta + \epsilon \tag{3.4}$$

A primary goal of regression analysis is to estimate the unknown parameters β. The standard approach is least squares regression, where the estimates chosen to minimize are,

$$\|y - X\beta\|_2^2 = \sum_{i=1}^{n}[y_i - (\beta_o + \beta_1 x_{1i} + \cdots + \beta_p x_{pi})]^2 \tag{3.5}$$

The normal equation which determines the minimizer of 3.5 is,

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{3.6}$$

Ridge Regression

Ridge Regression (RR) (Shewhart et al., 2015) is a method of estimating the coefficients of Multiple Linear Regression (MLR) models in scenarios where independent variables are highly correlated.

The theory was first introduced by Hoerl and Kennard in 1970 (Hoerl and Kennard, 1970). However, RR is a special case of Tikhonov regularization (Tikhonov, 1966), in which all parameters are equally regularized. The Penalized Residual Sum of Squares (PRSS) is,

$$\|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2 \tag{3.7}$$

RR estimate coefficients in 3.7 using

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y \tag{3.8}$$

Where λ is the shrinkage parameter and I, an identity matrix.

Neural Network

Neural Network (NN) is an interconnected group of nodes inspired by a simplification of neurons in the brain. The network learns by adjusting weights to reduce the prediction error(Han et al., 2011). Initially, all weights and biases are randomly allocated. The algorithm then runs iteratively, and each iteration comprises two steps: forward feeding and backpropagation (Phan, 2019).

- In the forward feeding phase, the output for the computation unit, a_i^{k+1} is the result of applying a transfer function, σ to the summation of all signals from each connection, a_i^k times the value of the connection weight, W_{ji} between node a_i^{k+1} and connection a_i^k (Coakley and Brown, 2000). The prediction of the output layer is then compared to the observed outcome to derive the learning rate and errors (Phan, 2019).

  $$z_j = \sum_j (W_{ji}^k a_i^k) \tag{3.9}$$

  $$a_i^{k+1} = \sigma(z_j) \tag{3.10}$$

  $$where \quad \sigma = ReLU(x) = max(0, x)$$

  where a is the activation of unit i in layer k, W_j is the weight of unit i in layer k and σ is the transfer function, Rectified Linear Unit (ReLU).

- In backpropagation, given the learning rate and errors, the network recalculates the weights and bias in hidden layers and makes appropriate changes to reduce prediction errors (Phan, 2019).

Chapter 4

DATA COLLECTION, ANALYSIS AND RESULTS

4.1 Original Dataset

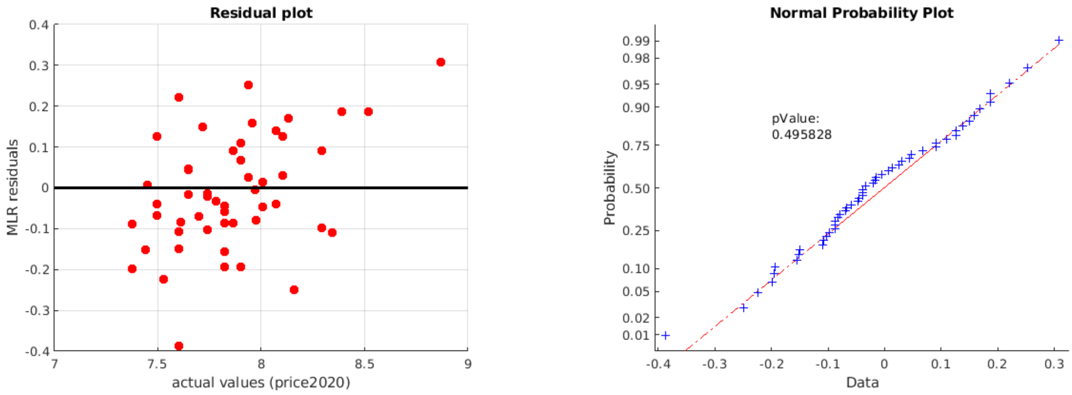
For this study, data obtained were through interviews with student residents in the hostels. The dataset is a cleaned dataset from 495 responses from 70 distinct hostels (Table 1). Since most hostels can be located in Ayeduase and Kotei, 38 and 20 hostels were selected from Ayeduase and Kotei, respectively, and 6 each from Bomso and Kentinkrono.

Table 4.1: Features Description

Name	**Type**	**Description**
location	categorical	general location of hostel
grade	numerical	average of students' evaluation
rank	categorical	overall quality of hostel
beds	numerical	beds in a room
study room	categorical	hostel's study room
tv room	categorical	hostel's tv room
security	categorical	security personnel or post
food joint	categorical	food joint within 5 minutes walk
external power	categorical	another source of power
ac	categorical	air conditioner in a room
proximity	numerical	distance to College of Science (CoS)
post code	categorical	post code of hostel
latitude	numerical	hostel's latitude
longitude	numerical	hostel's longitude
price2018	numerical	price of room for 2018/19 in cedis
price2019	numerical	price of room for 2019/20 in cedis
price2020	numerical	price of room for 2020/21 in cedis (target feature)

4.2 Results

First, we check the assumption of MLR model. Figure 4.1a shows no notable patterns in our residuals and a negligible correlation; hence, equality of variance and independence of observation assumptions are met. Figure 4.1b shows a plot with data points close to the normal line; a t-test on the residuals produced a p-value of >> 0.05; as a result, the normality assumption can be accepted. Since all assumptions are met, the MLR model qualifies as a candidate model for our dataset.



(a) Equality of Variance

(b) Normality of Residuals

Figure 4.1: Analysis of Residuals

The evaluation metrics given by equations 3.1, 3.2 and 3.3 of both train and test sets are presented in Table 4.2

Table 4.2: Evaluation Metrics of Train and Test sets

MODELS	TRAIN SET			TEST SET		
	MAE	RMSE	R^2	MAE	RMSE	R^2
MLR	0.1069	0.1337	0.7968	0.1114	0.1382	0.79
RR	0.1142	0.1406	0.7753	0.1116	0.1443	0.7711
NN	0.1069	0.1337	0.7968	0.1125	0.1398	0.7852

From Table 4.2, all three models performed well, with all coefficients of determination of the test set close to 1. However, the RR which was implemented to improve the fit in MLR utilizing its embedded feature selection properties with lambda of 0.0103, produced an R^2 lesser than the MLR model. Moreover, the MLR generalized well on the test set with an R^2 of 0.79; this could be as a result of no or negligible over-fitting and/or issues of collinearity.

The NN modeled the relationship between predictors and target features with two layers - a hidden layer with 10 nodes and an output layer with 1 node. The model, as noted by (Phan, 2019), works like a "black box," and we do not know the relationship between the predictors and the price prediction. Even so, the model performed exactly as MLR model on the train set. However, on the test set, there was a reduction in all three evaluation metrics (Figure 4.2).

Using the same models developed by our machine learning algorithms by learning from the entire dataset, we will explore the existence of submarkets (Maclennan and Tu, 1996) in the AYEDUASE and KOTEI areas using the table below.

Table 4.3: Evaluation Metrics of submarkets

AYEDUASE						
MODELS	TRAIN SET			TEST SET		
	MAE	RMSE	R^2	MAE	RMSE	R^2
MLR	0.1175	0.1487	0.6544	0.1253	0.1488	0.7786
RR	0.1194	0.15	0.6486	0.1327	0.1596	0.7451
NN	0.1175	0.1487	0.6544	0.1253	0.1488	0.7786

KOTEI						
MODELS	TRAIN SET			TEST SET		
	MAE	RMSE	R^2	MAE	RMSE	R^2
MLR	0.0913	0.1131	0.8686	0.1003	0.1188	0.8126
RR	0.0899	0.1192	0.8541	0.0854	0.1011	0.8642
NN	0.0913	0.1131	0.8686	0.1004	0.1188	0.8126

In Table 4.3 above, MLR and NN models produced similar results. However, the low number of observations may be a possible explanation for the poor performance of all models on the train set in the "AYEDUASE" submarket.

Chapter 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

In summary, this study is an exploratory attempt to use three machine learning algorithms to estimate hostel prices and then compare their performances after carefully preparing, transforming, and clearing the dataset. In this study, our models are trained with some hostel features and 2019/20 prices utilizing MLR, RR and NN. We have demonstrated the predictive power of machine learning algorithms on the hostel market, as evaluated by the performance metrics.

Given our dataset used in this paper, our main conclusion is that MLR and NN can generate comparably accurate hostel price estimations with lower prediction errors, compared with the RR results. Based on these estimates, we hope hostel managers, Students' Representative Council (SRC) and Private Hostels Association (PHA) will incorporate this approach in estimating hostel prices or developing new policies that govern the hostel market.

Furthermore, results from Table 4.3 prove the existence of a hostel submarket in the AYEDUASE and KOTEI areas, with all R^2 greater than 0.8 on the test sets.

5.2 Recommendation

First, the non-linear relationship between proximity and hostel price, which is mostly the main reason students rent hostels, should be explored using other algorithms such as random forest and decision trees.

Moreover, from our dataset, post_code contains 21 unique area codes. Further investigations into the hostel market should consider reducing the complexity of the feature by employing an unsupervised machine learning algorithm for cluster analysis to reduce the number of features during training.

Finally, this paper considered only the 2019/20 academic year information for the hostels. The time effect of the hostel price, which could potentially impact the estimated results, was removed due to the sparsity of such data. Therefore, we suggest that different data collection styles should be employed.

REFERENCES

C. Calhoun. Property valuation models and house price indexes for the provinces of Thailand: 1992-2000. *Housing Finance International*, 17(3):31, 2003. ISSN 1534-8784.

S. Chatterjee and J. S. Simonoff. *Handbook of Regression Analysis*. 2013. ISBN 9780470887165. doi: 10.1002/9781118532843.

J. R. Coakley and C. E. Brown. Artificial neural networks in accounting and finance: Modeling issues. *Intelligent Systems in Accounting, Finance & Management*, 9 (2):119–144, 2000.

M. J. J. Douglass. *Book Review: Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow, 2nd edition by Aurélien Géron*, volume 43. 2020. ISBN 9781492032649. doi: 10.1007/s13246-020-00913-z.

D. Friedland. A fresh look at data preparation. URL `https://www.iri.com/blog/business-intelligence/a-fresh-look-at-data-preparation/`.

E. K. Gavu and A. Owusu-Ansah. Empirical analysis of residential submarket conceptualisation in Ghana. *International Journal of Housing Markets and Analysis*, 12(4):763–787, 2019. ISSN 17538289. doi: 10.1108/IJHMA-10-2018-0080.

J. Han, J. Pei, and M. Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

A. Jaiantilal. Feature Selection by Iterative Reweighting: An Exploration of Algorithms for Linear Models and Random Forests. 2013.

N. Kamal, E. Chaturvedi, S. Gautam, and S. Bhalla. House Price Prediction Using Machine Learning. pages 799–811, 2021. doi: 10.1007/978-981-15-9774-9_73.

V. Limsombunchai. House Price Prediction : Hedonic Price Model vs . Artificial Neural Network Paper presented at the 2004 NZARES Conference House Price Prediction :. *New Zealand Agricultural and Resource Economics Society Conference*, pages 25–26, 2004.

R. J. Little and D. B. Rubin. *Statistical Analysis with Missing Data. Roderick J. A. Little, Donald B. Rubin*, volume 94. John Wiley & Sons, 1988. doi: 10.1086/228956. URL `https://b-ok.africa/book/1185981/a058d2`.

D. Maclennan and Y. Tu. Economic perspectives on the structure of local housing systems. *Housing studies*, 11(3):387–406, 1996.

C. H. Madhuri, G. Anuradha, and M. V. Pujitha. House Price Prediction Using Regression Techniques: A Comparative Study. *6th IEEE International Conference on Smart Structures and Systems, ICSSS 2019*, pages 4–8, 2019. doi: 10.1109/ICSSS.2019.8882834.

Mathworks. isoutlier. URL `https://www.mathworks.com/help/matlab/ref/isoutlier.html`.

K. Y. Ngiam and I. W. Khor. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5):e262–e273, 2019. ISSN 14745488. doi: 10.1016/S1470-2045(19)30149-4. URL `http://dx.doi.org/10.1016/S1470-2045(19)30149-4`.

N. Nguyen and A. Cripps. Predicting Housing Value: A Comparison of Multiple Regression Analysis and Ar tificial Neural Networks. *Journal of Real Estate Research,*, 22(3):313–336, 2001. ISSN 0041-0101.

A. Owusu-Ansah. Examination of the determinants of housing values in urban Ghana and implications for policy makers. *Journal of African Real Estate Research*, 2(1):58–85, 2012.

S. Pan and J. Zhong. Shuaidong Pan Jianyuan Zhong Abstract 2 Relevent Works 1 Introduction 3 Methodology. 2019.

A. Parashar. The 7 stages of preparing data for machine learning, June 2021. URL `https://pub.towardsai.net/the-7-stages-of-preparing-data-for-machine-learning-dfe454da960b`.

P. Patil. What is exploratory data analysis?, 2018. URL `https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15`.

T. D. Phan. Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia. *Proceedings - International Conference on Machine Learning and Data Engineering, iCMLDE 2018*, pages 8–13, 2019. doi: 10.1109/iCMLDE.2018.00017.

H. Selim. Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, 36(2 PART 2):2843–2852, 2009. ISSN 09574174. doi: 10.1016/j.eswa.2008.01.044. URL `http://dx.doi.org/10.1016/j.eswa.2008.01.044`.

W. A. Shewhart, S. S. Wilks, E. D. J. Balding, N. A. C. Cressie, and G. M. Fitzmaurice. *Paper Money Value Change: Comparative Banking Fiqhiyyah Study*, volume 5. 2015. ISBN 9781118644614. doi: 10.15408/aiq.v5i1.2561.

Y. Shweta Bhatt. 5 things you need to know about reinforcement learning, March 2018. URL `https://www.kdnuggets.com/2018/03/5-things-reinforcement-learning.html`.

M. Thamarai and S. P. Malarvizhi. House Price Prediction Modeling Using Machine Learning. *International Journal of Information Engineering and Electronic Business*, 12(2):15–20, 2020. ISSN 20749023. doi: 10.5815/ijieeb.2020.02.03.

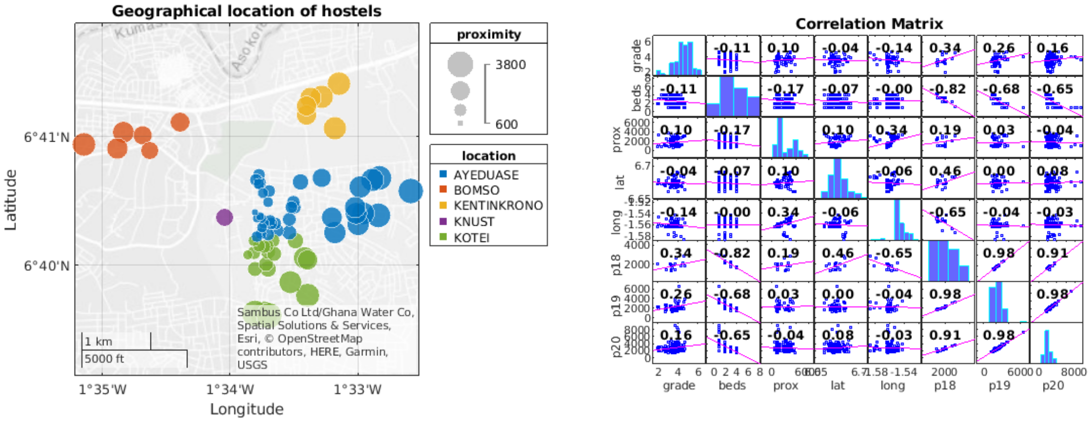
A. N. Tikhonov. On the stability of the functional optimization problem. *USSR Computational Mathematics and Mathematical Physics*, 6(4):28–33, 1966.

Q. Truong, M. Nguyen, H. Dang, and B. Mei. Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, 174(2019):433–442, 2020. ISSN 18770509. doi: 10.1016/j.procs.2020.06.111. URL `https://doi.org/10.1016/j.procs.2020.06.111`.

N. Vineeth, M. Ayyappa, and B. Bharathi. House Price Prediction Using Machine Learning Algorithms. *Communications in Computer and Information Science*, 837:425–433, 2018. ISSN 18650929. doi: 10.1007/978-981-13-1936-5_45.

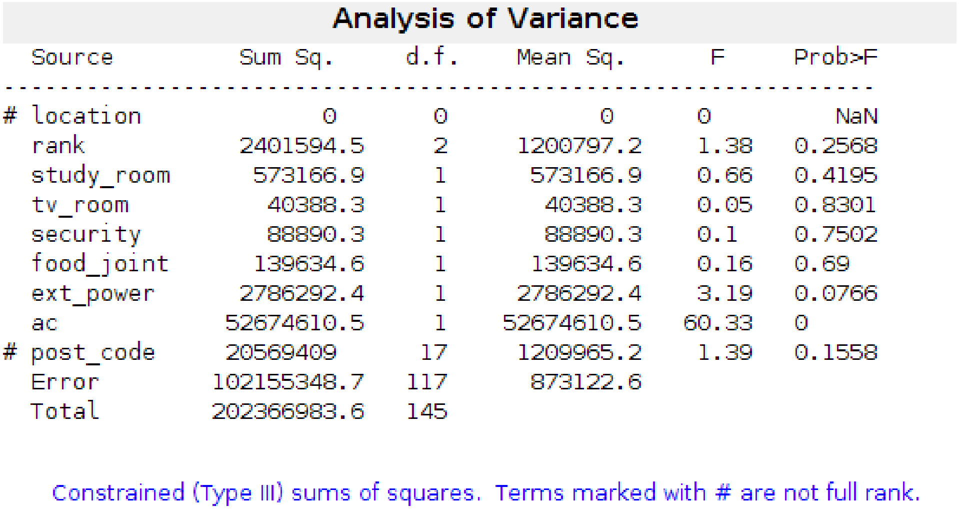
S. J. Xin and K. Khalid. Modelling House Price Using Ridge Regression and Lasso Regression. *International Journal of Engineering and Technology*, 7(4.30):498, 2018. doi: 10.14419/ijet.v7i4.30.22378.

APPENDIX A: Additional Informative Figures and Tables



(a) Geographical location of hostels

(b) Correlation matrix



```
                    Analysis of Variance
  Source          Sum Sq.     d.f.    Mean Sq.      F     Prob>F
-----------------------------------------------------------------
# location              0       0           0       0        NaN
  rank          2401594.5       2   1200797.2    1.38     0.2568
  study_room     573166.9       1    573166.9    0.66     0.4195
  tv_room         40388.3       1     40388.3    0.05     0.8301
  security        88890.3       1     88890.3    0.1      0.7502
  food_joint     139634.6       1    139634.6    0.16     0.69
  ext_power     2786292.4       1   2786292.4    3.19     0.0766
  ac           52674610.5       1  52674610.5   60.33     0
# post_code      20569409      17   1209965.2    1.39     0.1558
  Error       102155348.7     117    873122.6
  Total       202366983.6     145

    Constrained (Type III) sums of squares.  Terms marked with # are not full rank.
```

Figure 2: ANOVA table

Table 1: List of hostels and their location

hostel	**location**
ADOM - BI	AYEDUASE
AFRIM	AYEDUASE
AMANDAH	AYEDUASE
AMEN MAIN	AYEDUASE
AMERICAN HOUSE	AYEDUASE
ANAROSA	KOTEI
ANGLICAN	KENTINKRONO
B. O. EXECUTIVE	AYEDUASE
BANIVILLAS	KENTINKRONO
BEACON	AYEDUASE
BLUE ARK	BOMSO
BY HIS GRACE	AYEDUASE
CANAM	KOTEI
CASA MARIA	AYEDUASE
CELIA ROYAL	KOTEI
CHRISTIAN IPS	BOMSO
CRYSTAL ROSE	KENTINKRONO
DAKENS INTERNATIONAL	AYEDUASE
DELISA MAIN	AYEDUASE
DEVALYPAH	KOTEI
ENIN	AYEDUASE
FNF	AYEDUASE
F – PLAZA	AYEDUASE
FLINT	KOTEI
FOSUA HOMES	AYEDUASE
FRANCO	KOTEI
FRONTLINE INN	AYEDUASE
GAZA	KENTINKRONO
GEORGIA	KENTINKRONO
HALLOWED	KOTEI
HAPPY FAMILY	AYEDUASE
HIGH ACHIEVERS	AYEDUASE
HYDES	AYEDUASE
JALEX	AYEDUASE
JOHANNES	KOTEI

K GEE	AYEDUASE
KWAKYEWAA	KOTEI
LONG ISLAND	KOTEI
MANCHESTER	KOTEI
MASS	KOTEI
MILLENNIUM LIGHT	AYEDUASE
MORNING STAR PALACE	BOMSO
NANA ADOMAH	AYEDUASE
NEVADA	AYEDUASE
NO WEAPON	KOTEI
NYAME MIREKU	AYEDUASE
NYANTAKYI	AYEDUASE
ORANGE	KOTEI
P 3 HOSTEL	AYEDUASE
PINAMANG	AYEDUASE
PRESTIGE	KOTEI
PROVIDENCE	KOTEI
RISING STAR (NYBERG)	AYEDUASE
RISING SUN	AYEDUASE
ROYAL GATE	BOMSO
SHALOM KIBBUTZ	AYEDUASE
SHEPHERDSVILLE	AYEDUASE
SOMPA	KOTEI
SPLENDOR	AYEDUASE
STANDARD	BOMSO
SUN CITY	KENTINKRONO
THE BEST	KOTEI
THY KINGDOM COME	AYEDUASE
THY WILL BE DONE	KOTEI
ULTIMATE	BOMSO
VICTORY TOWERS	AYEDUASE
WAGYINGO	AYEDUASE
WEST END	AYEDUASE
WHITE HOUSE	AYEDUASE
WHITPAM A	KOTEI