Article

Impact of Green Features on Rental Value of Residential Properties: Evidence from South Africa

Tawakalitu Bisola Odubiyi [1], Rotimi Boluwatife Abidoye [2,*], Clinton Ohis Aigbavboa [3],
Wellington Didibhuku Thwala [4], Adeyemi Samuel Ademiloye [5] and Olalekan Shamsideen Oshodi [3]

[1] Department of Civil Systems Engineering, Technische Universität Berlin, 13555 Berlin, Germany;
odubiyi@tu-berlin.de

[2] School of Built Environment, University of New South Wales, Sydney, NSW 2052, Australia

[3] cidb Centre of Excellence, Faculty of Engineering and the Built Environment, University of Johannesburg,
Johannesburg 2028, South Africa; caigbavboa@uj.ac.za (C.O.A.); os.oshodi@gmail.com (O.S.O.)

[4] SARChl in Sustainable Construction Management and Leadership in the Built Environment, Faculty of
Engineering and the Built Environment, University of Johannesburg, Johannesburg 2028, South Africa;
didibhukut@uj.ac.za

[5] Department of Architecture, Caleb University, Lagos 11379, Nigeria; ademiloyeadeyemi30@gmail.com

[*] Correspondence: r.abidoye@unsw.edu.au; Tel.: +61-29065-7139

Abstract: In recent years, scholars have called for an increase in the usage of green features in the built environment to address climate change issues. Governments across the developed world are implementing legislation to support this increased uptake. However, little is known about how the inclusion of green features influences the rental value of residential properties located in developing countries. Data on 389 residential properties were extracted and collected from a webpage. Text mining and machine learning models were used to evaluate the impact of green features on the rental value of residential properties. The results indicated that floor area, number of bathrooms, and availability of furniture are the top three attributes affecting the rental value of residential properties. The random forest model generated better predictions when compared with other modelling techniques. It was also observed that green features are not the most common words mentioned in rental adverts for residential properties. The results suggest that green features add limited value to residential properties in South Africa. This finding suggests that there is a need for stakeholders to create and implement policies targeted at incentivising the inclusion of green features in existing and new residential properties in South Africa.

Keywords: classification model; data mining; green features; logistic regression; real estate; text mining

1. Introduction

The construction industry assembles the constructed spaces required for productive activities within an economy. Several years of research have shown that investments in infrastructure stimulate economic growth [1]. The products of the construction sector include buildings, highways, power plants and ports, among others. Despite its importance, activities in the construction sector generate about 40% of global greenhouse gas (GHG) emissions [2,3]. Also, greenhouse gases (such as carbon dioxide) and other air pollutants (such as PM_{10}), which are generated during the construction process, have adverse effects on human health and the environment. For instance, exposure to air pollutants has been linked to an increased risk of respiratory illness [4]. To address these problems, there have been calls for the adoption of sustainable practices in the construction sector.

The number of studies focused on identifying sustainable construction practices has been increasing. Research shows that the use of precast concrete components [5], a reduction in cement required for concrete through material replacement [6], the use of phase-change materials [7], and the inclusion of green areas into new developments [8] can reduce the

volume of GHGs attributed to construction works. Also, governments across the globe are implementing policies to encourage the use of green technologies. For instance, the government of Hong Kong and South Korea grants gross floor area concessions for property developments that incorporate green features [9]. The adoption of sustainable practices is beneficial to society and the environment. For instance, it would reduce emissions and improve air quality. Thus, there is a need to encourage stakeholders to embed green features in residential developments.

Researchers have reported that these green features are beneficial to the environment and occupants of buildings. For instance, gardens have been linked to improved food security, quality of life, productivity of people, and preservation of biodiversity within urban spaces [10–12]. Also, the use of technologies such as solar panels reduces the energy required by the occupants of properties [13]. In recent years, researchers have shown increased interest in examining the effect of green features on the value of real estate properties [14,15]. Voicu and Been [14] evaluated the effect of communal gardens on the value of real estate properties located in New York City. Several models have been developed for the prediction of the value of residential properties in developing countries [16,17]. However, the impact of embedding green features on the rental value of residential properties in developing countries remains unclear.

To address this gap in knowledge, data mining techniques were used to explore data and develop a predictive model for rental values of residential properties located in Cape Town, South Africa. The objectives of the study are (i) to examine the impact of green features on the rental value of residential properties and (ii) to assess the efficacy of using machine learning models for the prediction of rental values of residential properties. The data contribute to the existing knowledge in several ways. First, they provide insights into the effect of green features on the rental value of properties. Second, this study shows the efficacy of using machine learning models for the prediction of the rental value of properties. Third, the outcome of this study provides information to stakeholders for the development of strategies to encourage the inclusion of green features in residential properties. Finally, this study shows that the content of advertisements can be used to predict the value of real estate properties.

2. Materials and Methods

There is a large volume of research focused on modelling and predicting the value of residential properties. Existing research shows that structural features of real estate properties, the location of a building, and neighbourhood factors [18] are antecedents of the value of residential properties. The remainder of this section is structured into three subsections. The first section provides insights into the determinants of the value of residential properties. The second section highlights the effect of green features on the value of residential properties. The final section is a concise summary of methods used in previous research in predicting the value of residential properties.

2.1. Determinants of the Value of Residential Properties

Property values, the preferences of potential buyers, and investment decisions are determined by several factors. Over the years, research has shown that property values can be estimated as a function of these inherent determinants (features) that maximise utility for buyers and investors [19]. These attributes can be broadly grouped into structural, locational, and neighbourhood factors [18,20,21]. Based on evidence gleaned from the literature, it is evident that several factors (determinants) influence the value of residential properties.

Research into the determinants of residential property values has a long history. The determinants of the value of residential properties are summarised and presented in Table 1. The number of servants' quarters and number of bedrooms are the most important features influencing the value of residential buildings in Lagos, Nigeria [16]. In contrast, gardens have the highest impact on the value of residential properties in Hong Kong [22]. The variance in the Hong Kong and Nigerian studies could be linked to

the differences in the property market characteristics, government policies, and attitudes of local market participants.

Table 1. Summary of studies on residential property value determinants.

Category	Determinant	Source
Structural	Floor area	[16,17,23–25]
	Building age	[16,24]
	Number of bedrooms	[16,17,23–25]
	Number of bathrooms	[17,22–25]
	Swimming pool	[16,22,24]
	Garage	[24,26]
	Floor level	[27]
	Garden	[22,28]
	Balcony	[29]
	Green certification	[15]
	Lot size	[24]
	Fire place	[24]
	Air conditioning	[24]
Locational	Proximity to shopping malls	[25,30]
	Parking space	[16]
	Proximity to workplace	[31]
	Distance to school	[32]
	Distance to city centre/place of interest	[23,25,33]
	Proximity to transport infrastructure (such as train stations)	[25,34]
Neighbourhood	Urban green spaces	[20,25,35]
	Security	[25,36]
	Pollution levels	[25,37]

2.2. Impact of Green Features on Residential Property Values

Recently, more attention has focused on the effects of embedding green features on the value of residential properties. In the context of property economics, these studies were designed to provide empirical evidence to justify the inclusion of green features in the design of buildings. As mentioned in the introduction section of this paper, the inclusion of green features in buildings is beneficial to occupants and the environment. However, evidence showing economic benefits is essential for encouraging the inclusion of green features in residential buildings.

To date, several studies have investigated the effects of green features on the value of residential properties. The findings reported in previous research have been contradictory. Some studies have shown that urban green spaces (UGSs) have a positive effect on the value of residential properties [20,38]. In contrast, other studies [39,40] indicated that green features do not contribute to the value of residential properties. The variance in findings reported in previous research could be attributed to the following reasons: (i) proximity, (ii) characteristics of the neighbourhood, and (iii) preferences of tenants or investors in certain property markets. For example, [38] revealed that property prices tend to decrease as the distance to a UGS increases. Also, the socio-economic background of the local populace influences the acceptance of green roof systems [41]. The presence of communal gardens in poor neighbourhoods has been linked to crimes, anti-social behaviour, and offensive motifs or wall murals [20,36]. Based on the foregoing, it is clear that the effect of green features on the values of residential properties tends to vary based on the aforementioned three reasons.

The current study builds on existing literature by developing a model for predicting the rental value of residential properties. Also, the estimated model measures the effects of embedding green features on the rental value of residential properties in South Africa. To date, little is known about the economic value of embedding green features in residen-

tial buildings in the context of developing countries. This study fills the gap in current knowledge by providing insights into the impact of green features on the rental value of residential properties. The next section provides a summary of the literature on methods used for modelling and prediction of residential property values.

2.3. Methods Used for Predicting the Value of Residential Properties

Many methods have been used for modelling and prediction of the value of residential properties. A seminal study in this area is the work of [42], which utilised linear regression as a tool for modelling and prediction of the value of residential properties. The terms "regression model" and "Hedonic Price Models" (HPMs) are used interchangeably in the literature. Research has shown that the characteristics of data (such as non-linearity, multicollinearity, and outliers, among others) tend to affect the performance of regression models [43–45]. For example, [45] showed that neural network (NN) algorithms outperform regression models when applied to the prediction of property values. The predictive performance of the NN model indicates that the algorithm has the capability to capture the non-linearity present in real-world data. Thus, there has been a shift towards the use of non-linear models, such as data mining techniques, for the prediction of the value of properties.

The advancement in the field of computer science has led to the emergence of several data mining techniques. These new methods tend to address the limitations associated with the use of older techniques. For instance, local minima and overfitting are some of the weaknesses associated with NN models [46]. Newer methods, such as support vector machines, can overcome the limitations of the NN models. In recent years, newer methods (such as support vector machine and XGBoost) have been developed. Also, these newer methods generate better predictions when compared to older methods [47,48]. However, studies focused on evaluating the efficacy of some of these new methods are limited.

Various data mining techniques have been used for property valuation. The techniques found in the existing literature include support vector machine and decision tree, among others [22,27]. Previously, some of these data mining techniques were referred to as "black box" models, i.e., they provide no information on the effect of independent variables on the value of residential properties. However, this limitation has been addressed with the development of methods such as sensitivity analysis and variable importance [49]. Hence, the current study seeks to use seven data mining techniques for modelling and predicting the rental values of residential properties (see Section 3.4). The predictive performance of these methods is compared.

3. Methods

A multiplicity of methods was used to model the relationship between the attributes of a residential property and its value. Subsequently, the estimated model could be used for the prediction of the value of residential properties. The methods used for modelling the value of properties were discussed in the literature review section. The use of models offers an effective way to explore the relationship between property attributes and their economic value [50]. Two data mining techniques (text mining and logistic regression) were utilised in the current study. The coefficient of logistic regression provided insight into the strength of the relationship between the attributes of a residential property and its value.

Like previous research [16], the estimated model was validated by using it for out-of-sample prediction, and the data were collected from reliable secondary sources. This is a form of triangulation that improves the validity and credibility of the outcomes of research [51]. The combination of the quantitative findings from the logistic regression model and the qualitative findings from the text mining contribute to the rigour of the current study. Figure 1 shows the research framework utilised in this study.
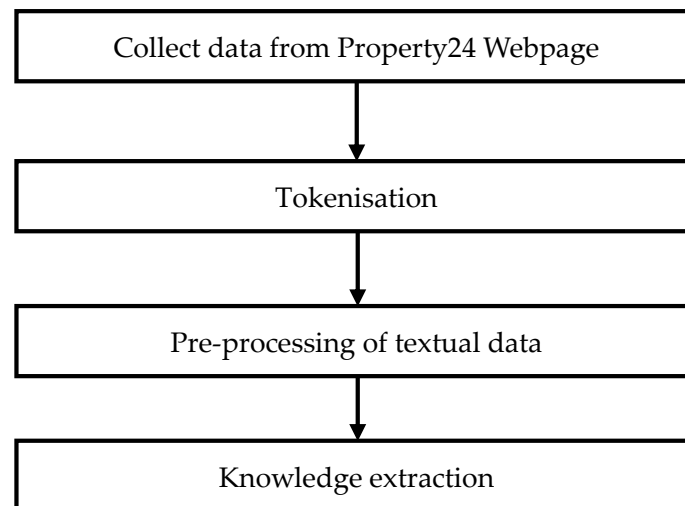
Figure 1. Research framework.

3.1. Data Collection

The data used in this study were collected from Cape Town, South Africa. Cape Town is one of the largest residential settlements in South Africa [52]. The city attracts many tourists due to its shoreline and beaches [52]. The city's seaport and a combination of public and private organisations provide employment and business opportunities to its residents [53]. Thus, the attraction of people to the city can be attributed to the opportunities that it provides.

The data used for this research were collected from the Property24 webpage (https://www.property24.com/). Property24 is one of the largest online platforms for advertising residential and commercial properties that are available for rent and sale. Each page contains information about the features of each property and its rental value. This webpage was adopted in similar previous studies on the value of real estate properties [17,54]. A total of 389 adverts were retrieved and used for analysis in this study. Because location has an impact on the value of residential properties, the data collected for this study were limited to three areas (i.e., City Centre, Sea Point, and Green Point) in Cape Town. These areas are within close proximity of Cape Town's City Centre. Each advert contains two types of information: numerical and text. The numerical data provide potential tenants with information on the property features, such as number of bedrooms and number of bathrooms, among others. The text data contain descriptive information about the residential property. For instance, text data provide information on the availability of furniture within the residential property.

The numerical and textual data for each residential property were collected. The data contained information on 13 independent variables which included the number of bedrooms, number of bathrooms, number of parking spaces, type of car parking facility (covered, open or street), dining room, lounge, availability of swimming pool, floor area, balcony, furniture (furnished or unfurnished), services (such as concierge), garden, and the distance between each advertised property and the nearest police station. The dependent variable was property rent. The numerical and textual data were analysed using several data mining techniques.

The statistics that describe the data collected for each residential property are presented in Table 2.

Table 2. Descriptive statistics relating to the collected data.

	Mean	Standard Deviation	Minimum	Maximum
Independent Variables				
Number of bedrooms	1.70	0.59	1	3
Number of bathrooms	1.43	0.51	0.5	3
Number of car park spaces	1.00	0.47	0	3
Number of dining rooms	0.37	0.49	0	2
Number of lounge/sitting rooms	0.85	0.55	0	3
Number of swimming pools	0.36	0.48	0	1
Floor area (m^2)	83.91	58.50	22	808
Distance to police station (Km)	1.68	3.27	0	32
Parking type	Covered (276), Open (57), Not available (56)			
Balcony	Yes (150) No (239)			
Furnished	Yes (214) No (175)			
Services	Yes (348) No (41)			
Garden	Yes (51) No (338)			
Dependent variable				
Property rent *	A (166) B (223)			

* This classification is based on income tax. See Section 3.2 and Table 3 for an explanation.

3.2. Data Processing

Textual and numerical data were collected from the webpage of "Property24". The text data were saved in txt files and numerical data were saved in .xls file format. In South Africa, the rent paid by tenants to a property owner or investor is classed as personal income [55]. The tax payable by the property owner is computed as a fraction of the annual rental income paid by tenants. In this study, the annual rents of the residential properties were computed by multiplying the monthly rental value by 12 months. Based on income tax brackets for South Africa [55], the residential properties were classified into two groups (A and B) using the annual rents payable by tenants to property owners (Table 3).

Table 3. Classification of rental price.

Personal Income Range	Income Tax Rate	Monthly Rental Price	Classification
ZAR 0–ZAR 189,880	18%	ZAR 1–ZAR 15,823	A
ZAR 189,881–ZAR 296,540	26%	ZAR 15,824–ZAR 24,711	B

Note: ZAR = South African Rands (15 ZAR = 1 USD).

3.3. Text Mining

Due to advances in computing, the volume of unstructured (textual) data has increased significantly. Traditionally, textual data are analysed using content analysis [56]. The growth in the quantity of textual data makes it difficult to use conventional methods to analyse these data. In recent years, text mining has emerged as an automated technique that can be used for the analysis of text. Text mining offers a robust approach to uncovering the underlying trends and patterns in text for making informed decisions. In previous studies, text mining was used to identify the determinants of hotel customers' satisfaction [57] and the vulnerability of software components [58], among others. In the current study, text mining was used to identify the features (determinants) of a residential property that influence its rental value. More specifically, text mining was used to extract and identify the determinants of the rental values of residential properties located in Cape Town.

A rising amount of research has shown that the content of an advert provides insight into the needs of the consumers and the value of the advertised product. For example, [59] found that the content of adverts influenced consumers' food choices. The findings presented in [60] indicate that the demand for a product is strongly influenced by the advertising content. It can be inferred that the content of adverts placed for residential

properties on the Property24 webpage is designed to provide information on attributes that meet the needs of potential tenants. Thus, the application of text mining to adverts for residential properties will provide insights into the factors influencing their values.

For the present study, a total of 389 adverts relating to residential property rental were collected from the Property24.com webpage. The process of using text mining for extracting relevant determinants of the rental value of residential properties is illustrated in Figure 1. R programming was used to implement the text mining phase of this study [61].

3.4. Machine Learning

As stated previously, text mining was used for the analysis of the collected textual data. In contrast, machine learning techniques were used for modelling numerical data. To assess the efficacy of using these methods for modelling and forecasting the rental prices of residential properties, seven techniques were applied in the current study. The techniques applied are neural network (NN), K-nearest neighbour (k-NN), logistic regression, boosted tree (XGBoost), random forest, and support vector machine. Additional details on the machine learning algorithms applied in the current study can be found in [62]. The logistic regression model was used as the benchmark for evaluating the predictive performance of other machine learning techniques.

Machine learning techniques can be applied to two types of predictive tasks, i.e., classification and regression. The classification task refers to the prediction of class labels. For example, [63] used machine learning models to predict the two classes, i.e., the chances that a student will drop out or not. For regression tasks, machine learning models are used for the prediction of numbers. Abidoye and Chan [45] used the neural network model for the prediction of the sales price of residential properties. In the current study, machine learning techniques were used for a classification task. The model predicted two class labels, i.e., class "A" or "B" (see Table 2).

The process of developing machine learning models can be divided into four distinct but interrelated phases. First, the independent variables to be included in the estimated models were identified. The outcome of the initial review of the literature and text mining provided justification for the variables used in model estimation. Second, the data used for the development of the models were collected from the Property24 webpage. Data relating to 389 residential properties were collected. Third, the machine learning models were trained. For the training process, the data were divided into two groups, i.e., training and test data sets.

The ratio of training data to test data used in previous research can be grouped into three clusters, i.e., 90:10, 80:20, and 70:30 [64]. In the current study, the collected data were randomly divided into two parts (70% training and 30% test data). To ensure that a robust model was identified, the cross-validation technique was used during the model training process. The test data set was used for model validation, i.e., to evaluate the ability of the model to predict unseen data. Finally (i.e., in the fourth stage), the trained machine learning models were used to predict the rental value of the residential properties in the test data set. The machine learning models were implemented using the R programming software [61] and the "tidymodels" package [65].

3.5. Methods for Evaluating Accuracy of Models

There are several metrics available for evaluating the predictive accuracy of classification models. However, the suitability of a particular metric is largely dependent on the circumstance of its application. Accuracy, sensitivity, specificity, recall, and precision are some examples of metrics used for evaluating models [66,67]. In this study, the trained models were evaluated using accuracy, F1 score, and Matthews Correlation Coefficient (MCC). The formula used for computing these metrics can be found in these published studies [67–69]. These metrics were adopted due to their usage in similar previous research [68,70]. The values of accuracy and F1 score range between 0 and 1 (a value close to 1 indicates that the model is generating reliable predictions). Also, the value of MCC

ranges between +1 and −1 (+1 refers to a perfect match between predicted and actual data; 0 indicates that the model is no better than a random flip of a coin; −1 represents a model in which there is no agreement between predicted and actual data).

4. Results

The results from the application of the two data mining techniques are presented in the subsequent subsections.

4.1. Text Mining

The files containing the textual data relating to the adverts were imported into the software. In the R programming environment, the text mining process was implemented. Based on the analysis of text, the most frequent words and co-occurrence of words are presented in the two subsequent sections (Figures 2 and 3).
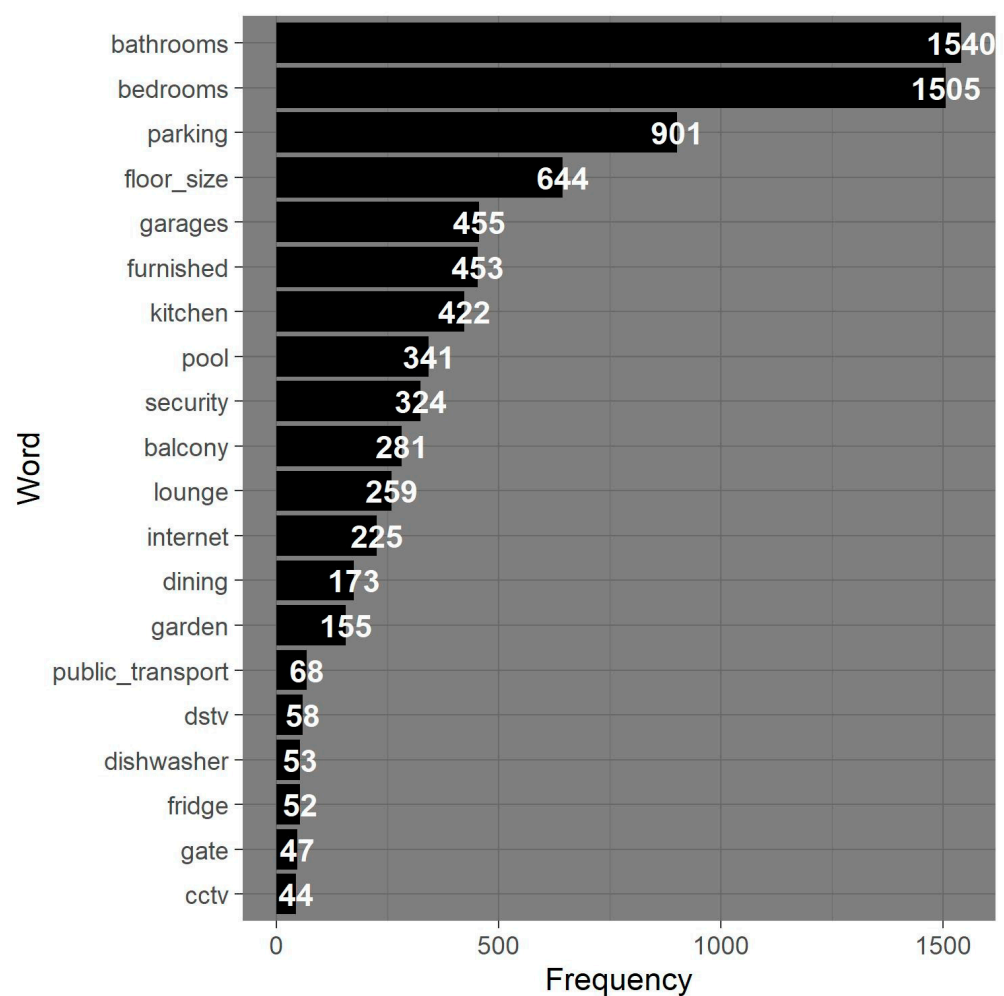

Figure 2. Most frequent 20 words in the adverts for residential properties.

4.1.1. Word Frequency

Figure 2 shows the 20 most frequent words in rental adverts for residential properties. Based on the data presented in Figure 2, it can be seen that "bathrooms" is the most frequent word mentioned in the rental adverts for residential properties. As expected, the number of bathrooms, number of bedrooms, parking, floor size, and furniture are the most mentioned words in the adverts for residential properties. The most frequently mentioned words give potential tenants insights into the size of spaces and the facilities available in the residential properties. Surprisingly, it was observed that sustainable features were not among the

most frequent words included in the rental adverts of residential properties, except for the "garden". This finding suggests that potential tenants are not willing to pay a premium for green features incorporated into residential properties.
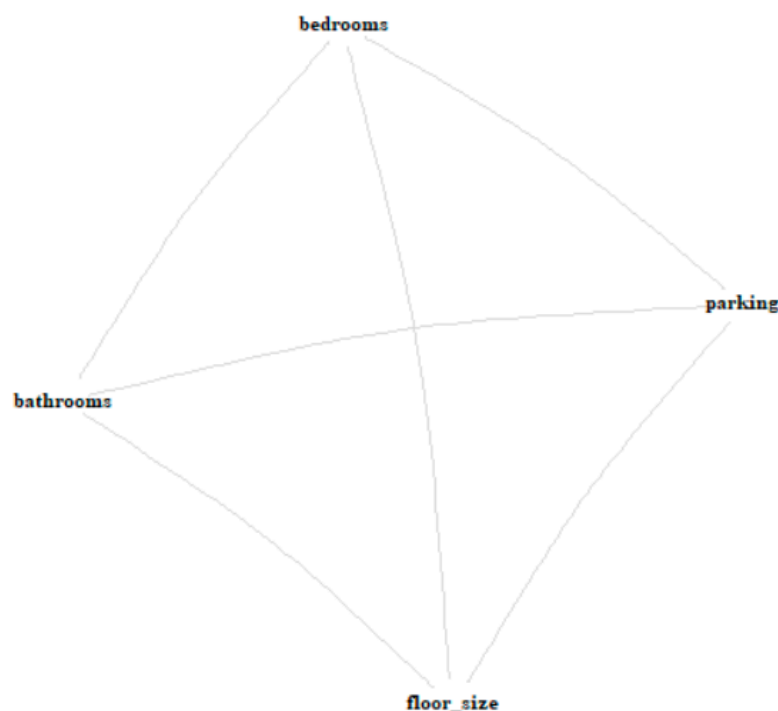
Figure 3. Word co-occurrence analysis of adverts for residential properties.

4.1.2. Word Co-Occurrence

Co-occurrence analysis was applied to the rental adverts of residential properties. In this phase, the co-occurrence analysis was used to identify the words that are mentioned frequently in the rental adverts in the residential properties. The results of word co-occurrence analysis are presented in Figure 3. As shown in Figure 3, it can be seen that the words "bathrooms", "bedrooms", "parking", and "floor-size" (floor area) were mentioned in all 389 rental adverts for residential properties. This finding suggests that "bathrooms", "bedrooms", "parking", and "floor-size" (floor area) are the most important determinants of the rental value of residential properties. In South Africa, inefficient public transport services have contributed to high levels of car ownership [71]. As expected, proximity to public transportation was rarely mentioned in the adverts for residential properties.

4.2. Predictive Performance of Machine Learning Models

To assess the efficacy of using machine learning models for the prediction of rental prices of residential properties, several models were trained, and the predictive performance of the models was compared. As stated previously, the collected data were divided into two groups: training (272/389—70%) and test sets (117/389—30%). The variables used to train the models are presented in Table 2. Table 3 shows the classification of rental prices that the models were trained to predict. The optimal parameters for the machine learning models were identified using grid search and 10-fold cross-validation of the training data. To prevent data leakage, only the training data were used for the process of training the models.

The optimal parameters of the machine learning models are shown in Table 4. The trained models were used to generate predictions of rental prices for the training data set. Also, the predictive performance of the developed models is summarised and presented in Table 4. For the in-sample evaluation of the trained models, two metrics (accuracy and F1 score) indicate that the neural network model generates a better forecast than other machine

learning models. However, the value of the MCC correlation coefficient indicates that the predictions from the support vector machine model are the most accurate, i.e., providing an improved level of agreement between the predicted and actual rental value of residential properties. This finding reiterates the importance of using multiple metrics to evaluate the performance of prediction models as suggested in [69].

**Table 4.** Predictive performance of machine learning models on training set.

Modelling Technique	Accuracy	F1 Score	MCC
Neural network	0.793	0.813	0.590
Support vector machine	0.779	0.909	0.779
Logistic regression	0.778	0.799	0.559
Random forest	0.770	0.796	0.557
XGBoost	0.769	0.794	0.546
K-nearest neighbour	0.767	0.779	0.541

The trained models were used to generate predictions for the test data set, i.e., the model validation process. Model validation is used to evaluate the ability of a model to predict previously unseen data. Using a similar approach, three metrics were used to evaluate the difference between the predictions and the actual rental value of residential properties in the test data set. Table 5 provides an overview of the error metrics emanating from the model validation process. The values of accuracy, F1 score, and MCC for the random forest model are closer to 1 when compared with those of the other model. This finding shows that the random forest model generates reliable predictions of the rental value of residential properties.

Table 5. Predictive performance of machine learning models on test set.

Modelling Technique	Accuracy	F1 score	MCC
Random forest	0.829	0.848	0.653
Logistic regression	0.794	0.810	0.595
Neural network	0.777	0.806	0.546
Support vector machine	0.777	0.797	0.556
XGBoost	0.735	0.760	0.467
K-nearest neighbour	0.735	0.763	0.463

Traditionally, machine learning models are referred to as "blackbox" models because the effect of the independent variables on the predicted variable is unknown. Sensitivity analysis has been proposed as one of the methods that can be used to uncover the strength of the relationship between variables incorporated into machine learning models [49]. Sensitivity analysis quantifies the contribution of each independent variable to improvement in the model's ability to make reliable predictions of rental value. Due to the predictive performance of the random forest model, sensitivity analysis was carried out on the model (see Figure 4). It can be seen from Figure 4 that floor area, number of bathrooms, and furniture are the best predictors of the rental value of residential properties. In contrast, gardens, pools, and dining areas had the least impact on the ability of the model to predict the rental value of residential properties.
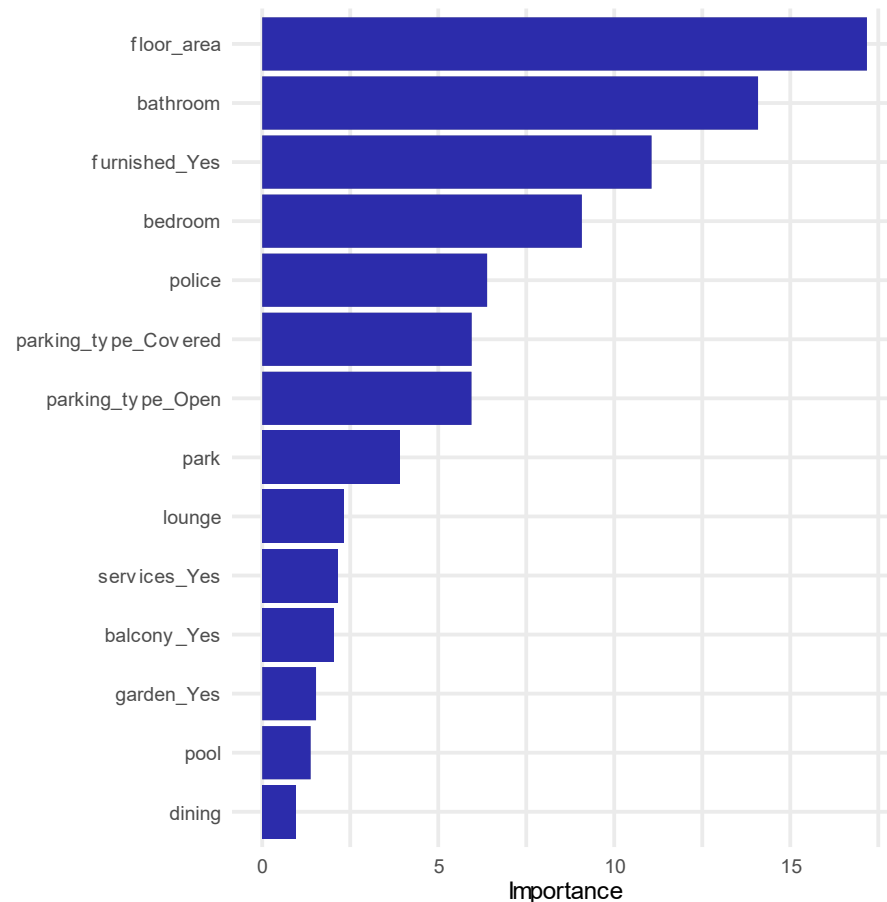
Figure 4. Sensitivity analysis of the random forest model.

5. Discussion

Very little was found in the existing literature on the impact of green building features on the rental values of residential properties. With respect to the gap in the current knowledge, the current study shows that gardens had little effect on the model's ability to predict the rental value of residential properties when compared with the floor area. Also, floor area, number of bathrooms, and availability of furniture were found to be the attributes that had the largest impact on the rental value of residential properties. Interestingly, the random forest model generates accurate predictions of the rental value of residential properties when compared to the other machine learning models that were developed in this study. Comparing the qualitative and quantitative data, it is evident that the frequently mentioned words in adverts are good predictors of the rental value of properties.

The results emanating from the current study are consistent with those reported in previous studies [16,72,73]. Abidoye and Chan [16] showed that the number of rooms in boys' quarters, number of bedrooms, sea view, and number of bathrooms are good predictors of the value of residential properties. Also, Nor et al. [72] found that size, location, and security have significant impacts on the rental value of residential properties. A study conducted in Germany showed that the age and size of a residential property have significant impacts on its rental value [73]. These results are consistent with those reported in Cespedes-Lopez's study [74], which showed that green features (such as green rating) do not influence the value of residential properties located in Alicante, Spain. Overall, the findings provide further support for the hypothesis that the size of residential properties is a good indicator of their value.

The findings reported here are contrary to previous studies, which have suggested that green features have a significant impact on the rental value of residential properties. For

instance, Choy et al. [28] showed that a positive relationship exists between the presence of a garden and the value of residential properties. Davis et al. [75] showed that improving the energy performance of a residential building results in an increase in its value. Gardens were rarely mentioned in adverts when compared to other significant features, such as the number of bedrooms. In addition, sustainable features, such as photovoltaic systems, are not the most commonly mentioned words in the rental advert for residential properties. The finding emanating from the current study suggests that green features have a limited impact on the rental value of residential properties when compared to floor area.

The non-convergence of the findings could be due to the uniqueness of the property market in various countries. For instance, Hong Kong has laws to encourage developers to incorporate green features into new residential developments. However, such laws do not exist in this study area. The low impact of green features suggests that there is little or no incentive for home owners and property developers to incorporate them into existing and new homes. One of the main issues that emerge from these findings is that there is a need for the development and implementation of policies targeted at encouraging the inclusion of sustainable features in residential properties.

6. Conclusions

An understanding of the effect of green features on the value of residential properties is a topical issue within the built environment. This information would provide a justification for the inclusion of green features in the designs of residential properties. This study has identified that floor area has a great impact on the ability to predict the rental value of residential properties. The impact of gardens on the rental value of residential properties is limited. This study has also shown that the estimated random forest model is a good tool for predicting the rental value of residential properties.

Based on the findings of this study, the following conclusions can be drawn: (i) there is a need to incentivise the inclusion of green features in residential properties; (ii) machine learning models are useful tools for the prediction of the value of residential properties; and (iii) the content of advertisements gives insights into the features that influence the value of a product. The predictive models reported here have extended our knowledge of the impact of size and other features of a residential property on its rental value. Although this study focuses on residential property valuation, the findings may be able to explain the underlying reasons for the slow uptake of green technologies in developing countries.

Being limited to Cape Town, the developed model may not be applicable to the whole of South Africa. The data used for the development of the models were secondary data (i.e., data published on the website). The authors did not check if there were differences between the listed price and finally agreed price. Notwithstanding this limitation, the triangulation of qualitative and quantitative data indicates that a robust approach was utilised in the present study. Advancements in the field of data science have led to the emergence of new methods. Further research should be undertaken in other cities in the developing world to validate or disprove the findings of the current study. Unless governments incentivise the inclusion of green features in residential properties in the developing world, sustainable development will not be achieved.

Author Contributions: Conceptualisation T.B.O., R.B.A. and O.S.O.; methodology, T.B.O., R.B.A. and O.S.O.; resources, C.O.A. and W.D.T.; writing—original draft preparation, T.B.O. and A.S.A.; writing—review and editing, R.B.A., C.O.A., W.D.T. and O.S.O.; visualisation, R.B.A. and A.S.A.; supervision, R.B.A., O.S.O. and W.D.T.; project administration, C.O.A., W.D.T. and O.S.O. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data for the study can be extracted from the website provided in the body of our paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dang, T.H.G.; Low, S.P. Role of construction in economic development: Review of key concepts in the past 40 years. *Habitat Int.* **2011**, *35*, 118–125.
2. Azzi, M.; Duc, H.; Ha, Q. Toward sustainable energy usage in the power generation and construction sectors—A case study of Australia. *Autom. Constr.* **2015**, *59*, 122–127. [CrossRef]
3. Liang, X.; Lin, S.; Bi, X.; Lu, E.; Li, Z. Chinese construction industry energy efficiency analysis with undesirable carbon emissions and construction waste outputs. *Environ. Sci. Pollut. Res.* **2021**, *28*, 15838–15852. [CrossRef] [PubMed]
4. Mebrahtu, T.F.; Santorelli, G.; Yang, T.C.; Wright, J.; Tate, J.; McEachan, R.R. The effects of exposure to NO_2, $PM_{2.5}$ and PM_{10} on health service attendances with respiratory illnesses: A time-series analysis. *Environ. Pollut.* **2023**, *333*, 122123. [CrossRef]
5. Ji, Y.; Li, K.; Liu, G.; Shrestha, A.; Jing, J. Comparing greenhouse gas emissions of precast in-situ and conventional construction methods. *J. Clean. Prod.* **2018**, *173*, 124–134. [CrossRef]
6. Ashish, D.K. Concrete made with waste marble powder and supplementary cementitious material for sustainable development. *J. Clean. Prod.* **2019**, *211*, 716–729. [CrossRef]
7. Al-Yasiri, Q.; Szabó, M. Incorporation of phase change materials into building envelope for thermal comfort and energy saving: A comprehensive analysis. *J. Build. Eng.* **2021**, *36*, 102122. [CrossRef]
8. Whittinghill, L.J.; Rowe, D.B.; Schutzki, R.; Cregg, B.M. Quantifying carbon sequestration of various green roof and ornamental landscape systems. *Landsc. Urban Plan.* **2014**, *123*, 41–48. [CrossRef]
9. Seo, B.; Kim, S.-H.; Lee, J.-H. Regulating Natural Lighting and Ventilation of Residential Buildings in Hong Kong Policy Implications for High-rise, High-density Housing Environments in South Korea. *Archit. Res.* **2014**, *16*, 81–92. [CrossRef]
10. Detweiler, M.B.; Murphy, P.F.; Myers, L.C.; Kim, K.Y. Does a wander garden influence inappropriate behaviors in dementia residents? *Am. J. Alzheimer's Dis. Other Dement.* **2008**, *23*, 31–45. [CrossRef]
11. Doody, B.J.; Sullivan, J.J.; Meurk, C.D.; Stewart, G.H.; Perkins, H.C. Urban realities: The contribution of residential gardens to the conservation of urban forest remnants. *Biodivers. Conserv.* **2010**, *19*, 1385–1400. [CrossRef]
12. Sellmann, D.; Bogner, F.X. Climate change education: Quantitatively assessing the impact of a botanical garden as an informal learning environment. *Environ. Educ. Res.* **2013**, *19*, 415–429. [CrossRef]
13. Kaya, O.; Klepacka, A.M.; Florkowski, W.J. Achieving renewable energy, climate, and air quality policy goals: Rural residential investment in solar panel. *J. Environ. Manag.* **2019**, *248*, 109309. [CrossRef] [PubMed]
14. Voicu, I.; Been, V. The effect of community gardens on neighboring property values. *Real Estate Econ.* **2008**, *36*, 241–283. [CrossRef]
15. Mesthrige, J.W.; Oladinrin, O.; Javed, A. Different grades and different green premiums: A cross sectional analysis of a green certification scheme. *Pac. Rim Prop. Res. J.* **2021**, *26*, 207–221. [CrossRef]
16. Abidoye, R.B.; Chan, A.P. Modelling property values in Nigeria using artificial neural network. *J. Prop. Res.* **2017**, *34*, 36–53. [CrossRef]
17. Oshodi, O.S.; Thwala, W.D.; Odubiyi, T.B.; Abidoye, R.B.; Aigbavboa, C.O. Using neural network model to estimate the rental price of residential properties. *J. Financ. Manag. Prop. Constr.* **2019**, *24*, 217–230. [CrossRef]
18. Chau, K.W.; Chin, T. A critical review of literature on the hedonic price model. *Int. J. Hous. Sci. Its Appl.* **2003**, *27*, 145–165.
19. McCord, M.; Davis, P.; Haran, M.; McGreal, S.; McIlhatton, D. Spatial variation as a determinant of house price: Incorporating a geographically weighted regression approach within the Belfast housing market. *J. Financ. Manag. Prop. Constr.* **2012**, *17*, 49–72. [CrossRef]
20. McCord, J.; McCord, M.; McCluskey, W.; Davis, P.; McIlhatton, D.; Haran, M. Effect of public green space on residential property values in Belfast metropolitan area. *J. Financ. Manag. Prop. Constr.* **2014**, *19*, 117–137. [CrossRef]
21. Abidoye, R.B.; Chan, A.P. Critical determinants of residential property value: Professionals' perspective. *J. Facil. Manag.* **2016**, *14*, 283–300. [CrossRef]
22. Lam, K.C.; Yu, C.; Lam, C.K. Support vector machine and entropy based decision support system for property valuation. *J. Prop. Res.* **2009**, *26*, 213–233. [CrossRef]
23. Lenaers, I.; Boudt, K.; De Moor, L. Predictability of Belgian residential real estate rents using tree-based ML models and IML techniques. *Int. J. Hous. Mark. Anal.* **2024**, *17*, 96–113. [CrossRef]
24. Sirmans, G.S.; MacDonald, L.; Macpherson, D.A.; Zietz, E.N. The value of housing characteristics: A meta analysis. *J. Real Estate Financ. Econ.* **2006**, *33*, 215–240. [CrossRef]
25. Gómez, K.; Iturra, V. How does air pollution affect housing rental prices in Chile? An economic assessment of $PM_{2.5}$ concentration across Chilean communes. *Environ. Dev. Econ.* **2021**, *26*, 364–380. [CrossRef]
26. Sirmans, S.; Macpherson, D.; Zietz, E. The composition of hedonic pricing models. *J. Real Estate Lit.* **2005**, *13*, 3–43. [CrossRef]
27. Fan, G.-Z.; Ong, S.E.; Koh, H.C. Determinants of house price: A decision tree approach. *Urban Stud.* **2006**, *43*, 2301–2315. [CrossRef]
28. Choy, L.H.; Mak, S.W.; Ho, W.K. Modeling Hong Kong real estate prices. *J. Hous. Built Environ.* **2007**, *22*, 359–368. [CrossRef]

29. Mesthrige, J.W.; Ka, C.W. Do buyers value a balcony as a green feature? An empirical analysis of the Hong Kong residential property market. *J. Real Estate Pract. Educ.* **2017**, *20*, 27–49. [CrossRef]

30. Zhang, L.; Zhou, J.; Hui, E.C.; Wen, H. The effects of a shopping mall on housing prices: A case study in Hangzhou. *Int. J. Strateg. Prop. Manag.* **2019**, *23*, 65–80. [CrossRef]

31. Yiu, Y.C. Housing price gradient changes between Macau and Hong Kong: A neighboring city effect. *Int. J. Hous. Mark. Anal.* **2008**, *1*, 195–206. [CrossRef]

32. Wen, H.; Xiao, Y.; Hui, E.C. Quantile effect of educational facilities on housing price: Do homebuyers of higher-priced housing pay more for educational resources? *Cities* **2019**, *90*, 100–112. [CrossRef]

33. Fauria, A.; Mathur, S. Impact of targeted redevelopment of central business district on housing prices in the surrounding neighborhoods: Evidence from Oakland, California. *J. Urban Plan. Dev.* **2012**, *138*, 244–253. [CrossRef]

34. Dubé, J.; Thériault, M.; Des Rosiers, F. Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992–2009. *Transp. Res. Part A Policy Pract.* **2013**, *54*, 49–66. [CrossRef]

35. Saló, A.; Gariga, A. The second-home rental market: A hedonic analysis of the effect of different characteristics and a high-market-share intermediary on price. *Tour. Econ.* **2011**, *17*, 1017–1033. [CrossRef]

36. Braakman, N. The link between crime risk and property prices in England and Wales: Evidence from street-level data. *Urban Stud.* **2017**, *54*, 1990–2007. [CrossRef]

37. Tian, G.; Wei, Y.D.; Li, H. Effects of accessibility and environmental health risk on housing prices: A case of Salt Lake County, Utah. *Appl. Geogr.* **2017**, *89*, 12–21. [CrossRef]

38. Liebelt, V.; Bartke, S.; Schwarz, N. Urban green spaces and housing prices: An alternative perspective. *Sustainability* **2019**, *11*, 3707. [CrossRef]

39. Couch, C.; Carswell, A.T.; Zahirovic-Herbert, V. An examination of the potential relationship between green status of multifamily properties and sale price. *Hous. Soc.* **2015**, *42*, 179–192. [CrossRef]

40. Rahman, F.; Rowlands, I.; Weber, O. Do green buildings capture higher market valuations and lower vacancy rates? A Canadian case study of LEED and BOMA-BEST properties. *Smart Sustain. Built Environ.* **2017**, *6*, 102–115.

41. Fernandez-Cañero, R.; Emilsson, T.; Fernandez-Barba, C.; Machuca, M.Á.H. Green roof systems: A study of public attitudes and preferences in southern Spain. *J. Environ. Manag.* **2013**, *128*, 106–115. [CrossRef]

42. Rosen, S. Hedonic prices and implicit markets: Product differentiation in pure competition. *J. Political Econ.* **1974**, *82*, 34–55. [CrossRef]

43. Worzala, E.; Lenk, M.; Silva, A. An exploration of neural networks and its application to real estate valuation. *J. Real Estate Res.* **1995**, *10*, 185–201. [CrossRef]

44. Lenk, M.M.; Worzala, E.M.; Silva, A. High-tech valuation: Should artificial neural networks bypass the human valuer? *J. Prop. Valuat. Investig.* **1997**, *15*, 8–26. [CrossRef]

45. Abidoye, R.B.; Chan, A.P. Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pac. Rim Prop. Res. J.* **2018**, *24*, 71–83. [CrossRef]

46. Bosman, A.S.; Engelbrecht, A.; Helbig, M. Visualising basins of attraction for the cross-entropy and the squared error neural network loss functions. *Neurocomputing* **2020**, *400*, 113–136. [CrossRef]

47. Lam, K.C.; Oshodi, O.S. Using univariate models for construction output forecasting: Comparing artificial intelligence and econometric techniques. *J. Manag. Eng.* **2016**, *32*, 04016021. [CrossRef]

48. Zendehboudi, A.; Baseer, M.A.; Saidur, R. Application of support vector machine models for forecasting solar and wind energy resources: A review. *J. Clean. Prod.* **2018**, *199*, 272–285. [CrossRef]

49. Cortez, P.; Embrechts, M.J. Using sensitivity analysis and visualization techniques to open black box data mining models. *Inf. Sci.* **2013**, *225*, 1–17. [CrossRef]

50. Fellows, R.; Liu, A. *Research Methods for Construction*, 4th ed.; Wiley Blackwell: Oxford, UK, 2015.

51. Love, P.E.D.; Holt, G.D.; Li, H. Triangulation in construction management research. *Eng. Constr. Archit. Manag.* **2002**, *9*, 294–303.

52. Lemanski, C. A new apartheid? The spatial implications of fear of crime in Cape Town, South Africa. *Environ. Urban.* **2004**, *16*, 101–112. [CrossRef]

53. Visser, G.; Kotze, N. The state and new-build gentrification in central Cape Town, South Africa. *Urban Stud.* **2008**, *45*, 2565–2593. [CrossRef]

54. Ghyoot, V. The accuracy of the Johannesburg provisional valuation roll, 2001. *Manag. Dyn. J. South. Afr. Inst. Manag. Sci.* **2008**, *17*, 2–13.

55. South African Revenue Service. Rates of Tax for Individuals. 2018. Available online: https://www.sars.gov.za/Tax-Rates/Income-Tax/Pages/Rates%20of%20Tax%20for%20Individuals.aspx (accessed on 1 June 2019).

56. Ignatow, G.; Mihalcea, R. *An Introduction to Text Mining: Research Design, Data Collection, and Analysis*; Sage Publications: Thousand Oaks, CA, USA, 2017.

57. Berezina, K.; Bilgihan, A.; Cobanoglu, C.; Okumus, F. Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *J. Hosp. Mark. Manag.* **2016**, *25*, 1–24. [CrossRef]

58. Scandariato, R.; Walden, J.; Hovsepyan, A.; Joosen, W. Predicting vulnerable software components via text mining. *IEEE Trans. Softw. Eng.* **2014**, *40*, 993–1006. [CrossRef]

59. Dixon, H.G.; Scully, M.L.; Wakefield, M.A.; White, V.M.; Crawford, D.A. The effects of television advertisements for junk food versus nutritious food on children's food attitudes and preferences. *Soc. Sci. Med.* **2007**, *65*, 1311–1323. [CrossRef] [PubMed]

60. Bertrand, M.; Karlan, D.; Mullainathan, S.; Shafir, E.; Zinman, J. What's advertising content worth? Evidence from a consumer credit marketing field experiment. *Q. J. Econ.* **2010**, *125*, 263–306. [CrossRef]

61. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.

62. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann: San Francisco, CA, USA, 2005.

63. Kemper, L.; Vorhoff, G.; Wigger, B.U. Predicting student dropout: A machine learning approach. *Eur. J. High. Educ.* **2020**, *10*, 28–47. [CrossRef]

64. Zhang, G.; Patuwo, B.E.; Hu, M.Y. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.* **1998**, *14*, 35–62. [CrossRef]

65. Kuhn, M.; Winkham, H. Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles. 2020. Available online: https://www.tidymodels.org (accessed on 10 June 2022).

66. Lever, J.; Krzywinski, M.; Altman, N. Classification evaluation. *Nat. Methods* **2016**, *13*, 603–604. [CrossRef]

67. Moccia, S.; De Momi, E.; El Hadji, S.; Mattos, L.S. Blood vessel segmentation algorithms—Review of methods, datasets and evaluation metrics. *Comput. Methods Programs Biomed.* **2018**, *158*, 71–91. [CrossRef] [PubMed]

68. Ullah, I.; Raza, B.; Malik, A.K.; Imran, M.; Islam, S.U.; Kim, S.W. A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access* **2019**, *7*, 60134–60149. [CrossRef]

69. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [CrossRef]

70. Zhou, J.; Li, X.; Mitri, H.S. Classification of rockburst in underground projects: Comparison of ten supervised learning methods. *J. Comput. Civ. Eng.* **2016**, *30*, 04016003. [CrossRef]

71. Luke, R. Car ownership perceptions and intentions amongst South African students. *J. Transp. Geogr.* **2018**, *66*, 135–143. [CrossRef]

72. Nor, M.I.; Masron, T.A.; Gedi, S.Y. Modeling of residential property rents in Somalia using two-stage modeling: Hedonic regression and artificial neural network. *Int. J. Hous. Mark. Anal.* **2019**, *13*, 331–356. [CrossRef]

73. Lorenz, F.; Willwersch, J.; Cajias, M.; Fuerst, F. Interpretable machine learning for real estate market analysis. *Real Estate Econ.* **2023**, *51*, 1178–1208. [CrossRef]

74. Cespedes-Lopez, M.F.; Mora-Garcia, R.T.; Perez-Sanchez, V.R.; Marti-Ciriquian, P. The influence of energy certification on housing sales prices in the province of Alicante (Spain). *Appl. Sci.* **2020**, *10*, 7129. [CrossRef]

75. Davis, P.T.; McCord, J.A.; McCord, M.; Haran, M. Modelling the effect of energy performance certificate rating on property value in the Belfast housing market. *Int. J. Hous. Mark. Anal.* **2015**, *8*, 292–317. [CrossRef]