International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Retail Site Selection using Machine Learning Algorithms

Hui-Jia Yee, Choo-Yee Ting, Chiung Ching Ho

Abstract*: Selecting a new site for retail business expansion has always been a challenge for decision-makers. It requires not only the sales data but the geographic data in order to decide the potential location for their respective purposes. Proper use of the data could lead to better decision-making. To date, common techniques such as geographic information system (GIS) and multi-criteria decision making (MCDM) have been applied to site selection. These methods, however, require not only extensive human effort but more importantly, difficult to validate the importance of identified variables. In this work, sales performance is proposed as a function of geospatial features to determine the suitability of a retail location. The main aim of this study was to identify features attributed to optimal site selection which in turn facilitate sales prediction for a telecommunication company in Malaysia. In this research, various feature selection techniques and machine learning models were deployed for sales prediction in order to determine the suitability of the new location. The findings show the top 3 feature selections are prediction step in VSURF, random search, and fuse learner with search strategy; the top 3 families are boosting, random forest and bagging; and the top 3 classifiers are C5.0, rf, and parRF. The crossover combination of the top feature selection-classifier can produce the AUC of more than 0.75. The highest AUC, 0.8354 was obtained through random search-parRF.

***Keywords*: Information Analytics, Machine Learning Comparison, Retail Site Selection.**

I. INTRODUCTION

Geospatial analytics is commonly known as location analytics, spatial analytics or spatial intelligence. It is often perceived as an intersection between business intelligence, geographic analysis, and data visualization [1], [2]. Up to the present, geospatial analytics has retained a vital role as solution to various domains including retail business [3], [4], [5], [6] energy conservation [7], [8], [9], [10] agriculture [11], safety planning [12], [13] and road network [14]. The targeted domain that is tackled in this paper is retail site selection. Retail site selection is an important concern related to the successfulness of a business. The decision making of the location of a business is a long-term investment as it is costly and hard to change [6]. In addition, a new establishment will bring financial and corporate image risks to the company [5]. However, the decision-making to locate a new retail store is expensive in terms of time, cost and human effort to investigate the prospect location.

In site selection, the conventionally, commonly and popularly used methods are multi-criteria decision making (MCDM) such as Analytic Hierarchy Process (AHP), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Fuzzy Set Theory and others. Besides, geographic information system (GIS) is also prevailing for gathering, analyzing, and visualizing the geographic data. MCDM identifies and evaluates a set of criteria and sub-criteria in decision making, while GIS overlays the data in different layers analyzes the spatial information on the map. However, in the past few years, machine learning has grabbed the attention of the public to solve various problems and extract insights and knowledge from the data. Researchers start to employ machine learning in site selection. The benefits of using machine learning in site selection are time-saving and the alternatives of location investigating can be much more a lot compare to using MCDM. There exist numerous machine learning techniques; the researchers tend to choose the best method in their opinion to apply in their problem. Anyhow, the optimal machine learning technique across all domains does not exist. Different machine learning approaches perform differently in different datasets and domains. Therefore, it is a great challenge for researchers to identify the optimal machine learning approach for solving the problem as every machine learning technique has its own strengths and weakness.

The objective of this study was to investigate the ability of machine learning in retail site selection other than the conventional methods. Secondly, this paper aims to identify the optimal machine learning family and the classifiers for retail site selection. As such, this paper evaluates over 42 classifiers from 13 families to find out which are the best suited for retail data to predict the sales performance of a telecommunication company in Malaysia and hence determine the location for a new establishment.

II. RELATED WORK

A. Retail Site Selection

Site selection is crucial in the retail business to ensure long-term success. Investors or franchisers would like to determine which prospective locations are highly potential to make good performance on the returns. Prior research shows the important role of location in retail site selection. [4] found that the presence of user attractors (e.g., train station and airport) and retail provisions of the same types to the target business are the factors that encode the location commercial competition of an area.

Manuscript published on November 30, 2019.
* Correspondence Author
Hui-Jia Yee*, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia. Email: huiijiia@gmail.com
Choo-Yee Ting, Institute of Postgraduate Studies, Multimedia University, Cyberjaya, Malaysia. Email: cyting@mmu.edu.my
Chiung Ching Ho, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia. Email: ccho@mmu.edu.my

The location attractors and types of retail stores are considered as the characteristics of a location. [4] and [15] use the popularity of an area as the indicator to identify the optimal location of a retail store. [15] utilized social media network data and with a set of features including business categories, locations, and neighbouring businesses to predict the popularity of an area.

[5] concluded that the features related to location and competition are most significant in affecting the performance of a retail business. In addition, [6] mentioned retail store site is important in providing a convenient location for customers. The following section discusses the various approaches by researchers in site selection.

B. Site Selection Techniques

Table I shows the techniques used by the researchers in site selection.

Previous research [16], [17], [11], [5], [9], [12], [6] employed Analytical Hierarchy Process (AHP) in site selection. Face-to-face interview [6], [5] and group decision-making process [12] that involve human judgment is used to choose the criteria and sub-criteria that will affect the alternative selection. After that, AHP is used to calculate the weights of each criterion and make a pairwise comparison between two alternatives. It is applied to a set of potential alternatives in order to select the best location for a defined objective.

Table- I: Site selection techniques

Author	AHP	GIS	BN	Multiple regression	TOPSIS	Monte Carlo	RST	Machine Learning
[16]	✓	✓						
[17]	✓	✓						
[11]	✓	✓						
[5]	✓	✓						
[9]	✓	✓						
[12]	✓	✓						
[6]	✓							
[18]	✓							
[19]		✓						
[3]		✓						
[8]		✓						
[13]		✓						
[20]			✓					
[21]			✓					
[7]				✓				
[22]					✓			
[23]						✓		
[24]							✓	
[15]								✓
[4]								✓

However, AHP becomes difficult mathematically to detect inconsistencies as the pairwise comparison between alternatives increases.

Another common approach in site selection is using geographic information system (GIS). In the retail business, [19] used GIS to develop a competitive location model to represent the market share and cannibalization effect on a map to locate a new hypermarket. In addition, [3] utilized GIS to produce consumption maps that focus on consumer expenses, which provide significant cues in determining the optimum location for the supermarket. In the energy field, [8] used GIS site-suitability analysis to identify the solar power site through the proximity raster layers that produced a map that shows the solar energy potential in the countries. For safety planning, [13] used GIS to model the topography of a hilly region for safe site selection. GIS as a visual inspection approach overlays different information on a map and allows analysis to perform on it. Nevertheless, human involvement is needed to conduct the exploratory analysis and derive hidden information from the layers [1]. Besides the commonly used AHP and GIS, other techniques such as Bayesian network (BN) [20], multiple regression [7], TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) [22], Monte Carlo simulation and optimization [23], rough set theory (RST) [24], and gradient boosting machine (GBM) [15] had also been employed in site selection.

C. Feature Selection

Machine learning has grabbed the attention of the public for acquiring valuable messages and information from the immense volume of data. However, when there is a fixed number of training data, the predictive power decrease when the dimensionality increase [25], and this is known as the Hughes phenomenon. The concept of feature selection is selecting a subset of the dataset by removing the redundant features in a way where the data still will be able to retain enough information for training purposes and improve the performance [26]. There are three general types of feature selection algorithms, namely filter, wrapper, and embedded methods. The filter method is independent of any machine learning algorithms. It uses statistical measures to assign a scoring to each variable and evaluates them. Then, the score is used to rank the variables. The advantages of filter methods are fast, scalable, do not run by a model and less computational complexity [27]. The examples of filter methods are the correlation, entropy, information gain [28], chi-square test, symmetrical uncertainty, and mutual information. Wrapper methods evaluate a subset of features by training a model on it. This method uses searching algorithms such as heuristic algorithm (forward and backward search), best first search, randomized hill-climbing, and greedy search to search for the best feature subset based on the model accuracy. In the searching process, different combinations of features are evaluated and compared to each other. The pros of wrapper methods are simple, clear and model feature dependencies. The examples of wrapper methods are recursive feature elimination, genetic algorithms, forward feature selection, and backward feature elimination. An embedded method is where the feature selection is built-in in the model. The similarity of the wrapper method and embedded method is to optimize the performance of the model. The most popular examples of the embedded method are LASSO and ridge regression, which have built-in penalization functions to reduce over-fitting [29].

Retrieval Number: D7186118419/2019©BEIESP
DOI:10.35940/ijrte.D7186.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The feature selection methods used in the research are: Boruta in Boruta package [30]; variable importance, recursive features elimination and genetic algorithm in caret package [31]; mutual information in entropy package [32]; best-first search, CFS, consistency, forward search, hill-climbing search, random forest filter (mean decrease in accuracy and mean decrease in node purity), Relief filter, chi-squared and OneR in FSelector [33]; fuse learner with search strategy, random search, and information gain in mlr [34]; conditional variable importance in party package [35], [36], [37]; superreduct subset computation in RoughSets [38]; and interpretation and prediction step in VSURF [39].

III. METHODOLOGY

The process flow of the study started with data preprocessing, data extraction, discretization, splitting, SMOTE and model training.

A. Dataset and data extraction

This section discusses the datasets and its source and methods used in this study. Table II shows the features extracted from different datasets and sources. The dataset used in this research was obtained from Telekom Malaysia (TM). The dataset included Point of Interest (POI), Yellow Pages (YP), iProperty, economy and education data. On the other hand, the demographic data were obtained from the Department of Statistics Malaysia (DOSM). The population was extracted from the local authority area, district and sub-district levels with the six races: Malay, other Bumiputera, Chinese, Indian, non-Malaysian citizens, and others. From POI and YP data, location features within 100m of an outlet were extracted. 100m was used as it is considered as the distance that a person is willing to walk from one place (shop) to another. The assumption was made to believe that the surrounded shops or businesses would influence the performance of an outlet. Table III showed the top 30 most frequently appear nearby location features for the telecommunication company. This set of location features might be different when extracted for other companies.

In addition, the property features were extracted within 1000m of the outlet from iProperty data. This will reveal the type of building, such as condominiums, storey shops, apartments, town house, etc. around the outlet. There are 26 types of property in the dataset.

Table- II: Features used in the study

Dataset/Source	Features	Number of features
Point of Interest and Yellow Pages	Nearby location features	30
iProperty	Nearby property	18
Economy data	Type of job, industrial field, employment status	34
Education data	Schooling status, education level, education certificate	22
Population	Race population	6
Google street view	Characteristic of outlets	6
Sales (Class label)	Yearly revenue	1

Table- III: Location features

Location features	
Chinese restaurants	Tour Operators
Fashion stores	Cake House & Dessert Beverages
Banks	Mamak
Malay restaurants	Computer & Telecommunication
Beauty Salon	Insurance Company
Café	Pharmacy
Private Clinic	Jewelry Shop
Building	Boutique / Beauty
Private Services Department	Hotel 2 Star
Insurance	Kurnia Agents
Lawyers	Private Dental Clinic
Restaurants	Hotel Budget
Shopping shop	Optical Shop
Associations	Private Clinic (PM Care)
Freight Forwarding	Post Office

Economic and education features were extracted at the district level. The economic data consists of the job types, industrial fields and employment status of the people in Malaysia. There are 9 job types, 21 industrial fields, and 4 different employment statuses recorded. On the other hand, education data represents the schooling status, educational level and education certificates of the people. There are 4 schooling status, 8 educational levels and 10 types of education certificates.

Besides, the characteristics features of the outlets were collected through Google Street View. The characteristic features include building type, visibility, public transport, parking space, type of access to the entrance and building type. The experts and the decision-makers of the company always considered these features when they plan to open an outlet. Hence, these features also include in the study.

B. Dataset aggregation and discretization

After extracted the features (location, property, economic, education, characteristic and population), the sales figure of the telecommunication company was aggregated to the features dataset to form the analytics dataset.

Therefore, the data structure of the analytics dataset includes the geospatial information as the predictor variables and the sales performance of the company as the response variables. An analytics dataset refers to tabular data that aggregate different dataset from various sources to form a data frame that fits analytics exercises [1].

There are 111 outlets for this telecommunication company. However, due to data insufficiency, there will be only 95 outlets included in this research. Since this is the classification problem, the sales performance is categorized into two classes, i.e. low sales performance, and high sales performance. There are 28 outlets with high sales performance and 67 outlets with low sales performance. The ratio of the classes is 0.29:0.71.

C. Splitting

One of the common approaches for machine learning is splitting the dataset into training and testing set. In this study, the dataset is split into 80% and 20%, where 80% was used for model training and 20% used for testing, i.e., to evaluate the performance of the model. The training and testing set were identical for all the models training to avoid the bias due to some training dataset perform better than the others. In the training process, 10-fold cross-validation was used for the classifiers to train the data. The experiment is repeated 5 times, that is using 5 different randomly split training and testing set to obtain the final averaged result to avoid biases of certain splitting seed.

Retrieval Number: D7186118419/2019©BEIESP
DOI:10.35940/ijrte.D7186.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

D. SMOTE

The total observations of the dataset are small, and the class is imbalanced, hence the SMOTE process is performed. SMOTE, known as Synthetic Minority Over-Sampling Technique is used to deal with the unbalanced classification problems [40]. It can generate a new SMOTE-d dataset with balanced classes [41] at the same time increase the observations of the dataset. In the present study, SMOTE is performed through the DMwR package [42] in R. R is a language and software environment for data manipulation, calculation, and graphical displays. It has been widely used in statistics, data mining, and machine learning.

The implementation of this study is using R through various packages and their functions. For instance, SMOTE function in the DMwR package is used to SMOTE the data. The parameter perc.over and perc.under of the function SMOTE are set to 500 and 120 respectively to generate five extra cases from the minority class and select 1.2% of the extra cases from the majority class for each case generated from the minority class. After the dataset is SMOTEd, the total observations become 276 with 138 high sales performance and 138 low sales performances. Now, the ratio of the classes is 0.5:0.5, the class is balanced. The geospatial details include nearby establishment and property, demographics, economic condition and education status of the residents around the outlet. There is a total of 116 features extracted in the dataset. Due to the high dimensionality of feature space, feature selection will be employed to eliminate the Hughes phenomenon. This work has implemented 20 feature selection methods to obtain the feature subsets for subsequent classification. There will be 20 datasets with respective features selected to evaluate the classifiers. Table IV shows the list of classifiers and their families that will be used in this study. This work implemented the classifiers in the caret package as this package consists of miscellaneous functions for classification and regression models.

E. Performance metrics

In the experimental work, 38 classifiers from 13 families were evaluated over 20 feature subsets. The area under the curve (AUC) is used to evaluate the performance of each classifier. Receiver Operating Characteristics (ROC) graph is used to organize classifiers and visualize the performance of the classifiers (Fawcett, 2006). ROC graph is plot based on the true positive rate (tpr) and false positive rate (fpr). True positive rate is also known as recall or sensitivity. It measures the percentage of correctly classified cases in positive outcomes. The equation of tpr is defined as:

$$tpr = \frac{tp}{tp + fn}$$

where the number of positive cases being classified correctly (tp) over the total number of positive cases ($tp + fn$). False positive rate (fpr) is the probability of falsely predicting the negative outcome as positive. It is mathematically expressed as:

$$fpr = \frac{fp}{fp + tn}$$

Table- IV: Classifiers and their families

Family	Classifier	Abbreviation
Discriminant Analysis (DA)	Linear Discriminant Analysis	lda2
	Shrinkage Discriminant Analysis	sda
	Linear Discriminant Analysis with Stepwise Feature Selection	stepLDA
	Quadratic Discriminant Analysis with Stepwise Feature Selection	stepQDA
	Flexible Discriminant Analysis	fda
	Penalized Discriminant Analysis	pda
Neural Networks (NN)	Radial Basis Function Network	rbfDDA
	Multi-Layer Perceptron	mlp
	Model Averaged Neural Network	avNNet
	Multi-Layer Perceptron (Weight Decay)	mlpWeightDecay
	Neural Network	nnet
	Neural Network with Feature Extraction	pcaNNet
Decision Trees (DT)	CART (Classification and Regreesion Trees) (complexity parameter)	rpart
	CART (max tree depth)	rpart2
	Single C5.0 tree	C5.0Tree
	Conditional Inference Tree (tunes over mincriterion)	ctree
	Conditional Inference Tree (tunes over maxdepth)	ctree2
	C4.5-like Trees	J48
Rule-Based methods (RL)	Rule-Based Classifier	PART
	Single C5.0 Ruleset	C5.0Rules
	Rule-Based Classifier	JRip
	OneR	OneR
Boosting (BST)	C5.0	C5.0
	eXtreme Gradient Boosting (Tree)	xgbTree
	Extreme Gradient Boosting (Linear)	xgbLinear
Bagging (BAG)	Bagged CART	treebag
Random Forests (RF)	Random Forest	rf
	Regularized Random Forest	RRF
	Parallel Random Forest	parRF
	Regularized Random Forest (global)	RRFglobal
Generalized Linear Models (GLM)	Bayesian Generalized Linear Model	bayesglm
	Generalized Linear Model with Stepwise Feature Selection	glmStepAIC
Nearest Neighbor methods (NN)	k-Nearest Neighbors	knn
Partial Least Squares and Principal Component Regression (PLSR)	Partial Least Squares	pls
Logistic and Multinomial Regression (LMR)	Penalized Multinomial Regression	multinom
Multivariate Adaptive Regression Splines (MARS)	Multivariate Adaptive Regression Splines	gcvEarth
Other Methods (OM)	Nearest Shrunken Centroids	pam
	Gaussian Process with Radial Basis Function Kernel	gaussprRadial

where it is calculated as the number of negative cases being predicted as positive (fp) over the total number of negative cases ($fp + tn$).

ROC plots tpr against fpr at different classification thresholds as shown in Figure 1.

Retrieval Number: D7186118419/2019©BEIESP
DOI:10.35940/ijrte.D7186.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

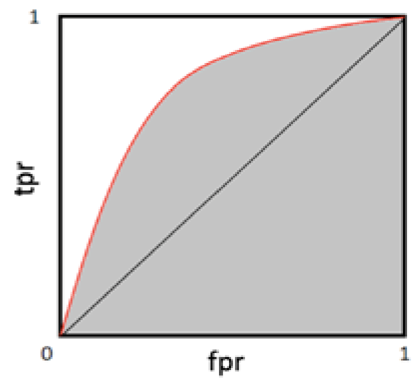


Fig. 1. ROC curve

The shaded region in the figure is the area under the ROC curve (AUC), which ranges from 0 to 1. The higher the AUC, the better the classification performance. The diagonal in the graph indicates a random classifier that has an AUC of 0.5. As tpr increase, fpr will increase as well. Hence, the closer the ROC curve is to the upper left corner, i.e. point (0,1), the better the trade-off between tpr and fpr, and the higher the area under the ROC curve, and thus the better the classifier [43].

IV. RESULT AND DISCUSSION

A. Discussion by machine learning family and classifiers

The AUC for the classifiers and their families respectively were shown in Table V. The highest AUC is bolded and underlined. As shown in Table V, according to the AUC, boosting (BST) family achieves the highest AUC (0.7417), the second-highest family is random forests (RF), AUC = 0.7352 follows by bagging (BAG) (AUC = 0.7239).

Boosting refers to a family of algorithms that converts weak learners to strong learner [44]. The process of boosting is as follows: firstly, the first base learner assigns equal weight to the inputs or observations. It will fit a subsample of the data and make a prediction. Any misclassified observation will be given a higher weight. The following learner will pay attention to the incorrectly predicted observations. Boosting is an iterative process; it will continue to add classifier learners until it reaches the limit of the number of the model or higher accuracy is achieved. Boosting builds a strong learner from the combination of the outputs of the weak learners that eventually improve the predictive accuracy of the model.

Random forest is a supervised machine learning algorithm, it creates the forest with several decision trees. The more trees in the forest, the more robust the model, the higher the accuracy. Given a training set with size s, a sample is taken from the training set with replacement. If there are P variables or features, a number $p < P$ features are selected. The best split on p is used to split the node. As the forest grows, the value of p remains constant. Each tree can grow to its largest extent possible and there is no pruning. By aggregating the result of the trees (i.e. voting for classification and average for regression), the prediction on new data is made. RF can be used in classification and regression problems. It can handle the missing value, large datasets with high dimensionality, do not overfit the model and measure the importance of the variables.

Table- V: AUC for classifiers and families

Family	Classifier	AUC for classifier	AUC for family
DA	lda2	0.6194	0.6282
	sda	0.6091	
	stepLDA	0.5900	
	stepQDA	0.6267	
	fda	0.7023	
	pda	0.6218	
NNET	rbfDDA	0.6710	0.6255
	mlp	0.5392	
	avNNet	0.6895	
	mlpWeightDecay	0.5390	
	nnet	0.6323	
	pcaNNet	0.6823	
DT	rpart	0.6395	0.6753
	rpart2	0.6860	
	C5.0Tree	0.7087	
	ctree	0.6643	
	ctree2	0.6479	
	J48	0.7051	
RL	PART	0.6891	0.6573
	C5.0Rules	0.6840	
	JRip	0.6756	
	OneR	0.5805	
BST	C5.0	**0.7542**	**0.7417**
	xgbTree	0.7405	
	xgbLinear	0.7305	
BAG	treebag	0.7239	0.7239
RF	rf	0.7436	0.7352
	RRF	0.7181	
	parRF	0.7427	
	RRFglobal	0.7364	
GLM	bayesglm	0.6433	0.6290
	glmStepAIC	0.6146	
NN	knn	0.6755	0.6755
PLSR	pls	0.6286	0.6286
LMR	multinom	0.6224	0.6224
MARS	gcvEarth	0.6776	0.6776
OM	pam	0.6387	0.6828
	gaussprRadial	0.7268	

Bagging is also known as bootstrap aggregating; it is a method to improve the predictive power or reduce the error rate of classifier learning algorithms. In the bagging procedure, given training set with s size, bagging will sample a new training set with replacement from the training set, hence the new training set will have the same size s with the original training set [45]. The new training set is expected to have 63% of the original examples in the bootstrap samples. Subsequently, a learning algorithm is applied to the training set and the procedure is repeated m times. In other words, m training set is sampled, and m model is fitted into the bootstrap samples. The result of the procedure is obtained by averaging the aggregated outputs (for regression) or voting (for classification). Breiman, the author of bagging suggested that, for the bagging to work, the learners must be unstable, that is a small change in the training set can cause significant changes in the model. Examples of unstable learning algorithms are decision trees and neural networks.

Retrieval Number: D7186118419/2019©BEIESP
DOI:10.35940/ijrte.D7186.118419
Journal Website: www.ijrte.org

On the other hand, stable learner such as nearest neighbor has little value when apply bagging to them or worse degrade the performance. Overall in this experiment, *C5.0* from the BST family is the classifier that performs best; it obtained an average AUC of 0.7542 across the 20 feature subsets. *C5.0* can achieve the maximum AUC of 0.8185 on the feature subset produced by fuse learner with search strategy. The lowest AUC is 0.6662 which is when it applied to the feature subset retained by the sequential search. *C5.0* in the caret package created a boosting ensemble of C5.0 decision trees and rule models, with and without winnow (feature selection), tuning the number of boosting trials in {1, 10, 20} [46]. The second best-performed classifier is *rf* (AUC = 0.7436) from the RF family, the maximum AUC is 0.8292 and the minimum AUC is 0.6369. For *rf*, the maximum AUC is obtained from the *random search* feature subset, whereas the minimum AUC is obtained from both feature subset produced by best first search and greedy search feature selection algorithm. *rf* creates a random forest with *ntree* = 500 and tuning the parameter

mtry with value 2:3:29. *ntree* is the number of trees to grow and *mtry* is the number of variables randomly sampled as candidates at each split. Subsequently, the third-best classifier is *parRF* (AUC = 0.7427) from the RF family too. *parRF* uses a parallel implementation of random forest with *mtry* = 2:2:8. Similarly, the maximum AUC (0.8354) is obtained from the random search feature subset and the minimum AUC (0.6769) is obtained from the best first search and greedy search feature subset. The maximum AUC obtained by *parRF*, 0.8354 is also the highest AUC achieved in this experiment.

There is a total of 8 classifiers from the top 3 best families. All the classifiers from the top 3 families outperformed other classifiers and they are in the top 10 best classifiers among all the classifiers employed in the experiment.

B. Discussion by feature selection

Table VI shows the average number of features retained by each of the feature selection methods in the 5 repeated experiments. Each feature subset will be trained by the 38 classifiers. The average AUC is obtained by averaging the AUC of the 38 classifiers across the feature subset.

Among the 20 feature selection methods, the prediction step in VSURF shows the highest AUC (0.7) follow by random search (AUC = 0.6883) and fuse learner with search strategy (AUC = 0.6869). The average number of features retained across the 5 experiments by prediction step in VUSRF is 6, by random search is 59 and by fuse learner with search strategy is 55.

Table VII shows the features retained by the top 3 feature selection methods. The feature retained marked with a tick is the feature selected at least one time by the feature selection method and the bolded and underlined tick denotes the feature selected are always (5 times) selected by the feature selection method. In prediction step in VSURF, Chinese and Malay are always selected; in random search, the features always being selected are professional, scientific and technical activities, offshore bodies and organizations, technician and associate professionals, and skilled works or carpenters; while in fuse learner with search strategy, the always selected features are Malay restaurants, still schooling, household activity as an employer, others (race), low secondary education and one storey shop house. In addition, there are 16 features that selected at least once in all the three feature selection methods, i.e. Malay restaurants, tour operators, one storey shop house, Malay, Chinese, others (race), skilled works or carpenters, basic jobs, electricity, gas, steam and air conditioning, administrative and support services activities, family workers

Table- VI: AUC for feature selection

Package	Method	Number of features retained	AUC
Boruta	Boruta	79	0.6750
caret	genetic algorithm (GA)	36	0.6685
caret	recursive features elimination (RFE)	17	0.6745
caret	variable importance	20	0.6517
entropy	mutual information (MI)	20	0.6523
Fselector	best first search (BFS)	6	0.6655
Fselector	greedy forward search (GFS)	6	0.6655
Fselector	hill climbing search (HCS)	55	0.6357
Fselector	CFS	16	0.6603
Fselector	consistency	7	0.6521
Fselector	Relief	20	0.6266
Fselector	random forest (mean decrease in node purity)	20	0.6684
Fselector	random forest (mean decrease in accuracy)	20	0.6774
mlr	Sequential search (SS)	2	0.6501
mlr	fuse learner with search strategy (fuse)	55	0.6869
mlr	random search (RS)	59	0.6883
party	conditional variable importance	20	0.6678
roughsets	superreduct subset computation	10	0.6862
VSURF	interpretation step	16	0.6740
VSURF	prediction step (pred)	6	**0.7000**

without salary, preschool education, pre-university, no certificate, Malaysian higher school certificate or equivalent, type of area (city or town). This shows that these attributes play an important role in influent the prediction.

In this study, the AUC obtained through the combination of the top feature selection, prediction step in VSURF and the top classifier C5.0 is 0.8092. However, this is not the highest AUC obtained from the experiment. The highest AUC, 0.8354 is obtained through the combination of random search (second-best feature selection) and *parRF* (third-best classifier). Besides, the cross-over combination of the top 3 feature selection with the top 3 classifiers can achieve the AUC of more than 0.75.

C. Discussion by time taken

Besides AUC, the time taken is discussed in this section. The training time of a classifier is recorded.

Table VIII shows the average time taken by each family and classifier to complete the experiments across all the feature subsets. BST shows the longest time taken (16.35s) to run the experiment. The second slowest family is GLM, which took 13.86 s to finish the experiment and the third slowest family is RF (6.03s). As shown in the figure, the rest families took less than 3 seconds to train the model. The result has shown that machine learning is a cost-saving method for retail site selection. With relevant data, the classifiers are able to predict the sales performance of a given location within minutes.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

This is not only saved in time but also the labour effort to identify the suitable retail site. However, there is more than one classifier in those slower families, and there is a possibility that the longer time taken is caused by a certain classifier in the family but not all the classifiers in the family.

Table- VII: The features retained by the top three feature selection methods

Features	pred	RS	fuse
Chinese restaurants		✓	✓
Fashion stores		✓	✓
Banks		✓	✓
Malay restaurants	✓	✓	✓
Beauty saloon		✓	✓
Cafe			✓
Private clinic		✓	
Building		✓	✓
Private services department		✓	✓
Insurance		✓	✓
Lawyers		✓	✓
Restaurants		✓	✓
Shopping shop		✓	✓
Associations		✓	✓
Freight forwarding		✓	✓
Tour operators	✓	✓	✓
Cake house & dessert beverages		✓	✓
Mamak		✓	✓
Computer and telecommunication shop		✓	✓
Insurance company		✓	✓
Pharmacy		✓	✓
Jewelry shop		✓	✓
Boutique/Beauty		✓	✓
2 Star hotel		✓	✓
Kurnia Agents		✓	✓
Private dental clinic		✓	✓
Budget hotel		✓	✓
Optical shop		✓	✓
Private clinic (PM Care)		✓	✓
Post office		✓	✓
Detached house		✓	✓
One storey shop house	✓	✓	✓
Two storey shop house	✓		✓
Three storey shop house			✓
Four storey shop house		✓	✓
Semi-detached house			✓
Low-cost house		✓	✓
Squatter house		✓	✓
Kampong house		✓	✓
Low-cost flat		✓	✓
Bungalow		✓	✓
Condominiums		✓	✓
Apartments		✓	✓
Chalet		✓	✓
Town house		✓	✓
One storey shop		✓	✓
Four storey shop		✓	✓
Five storey shop		✓	✓
Malay	✓	✓	✓
Other Bumiputera		✓	✓
Chinese	✓	✓	✓
Indians		✓	✓
Others (race)	✓	✓	✓
Non-Malaysian	✓	✓	
Manager		✓	✓
Professional		✓	✓
Technician and associate professionals		✓	✓
Clerical support workers		✓	✓
Service and sales workers		✓	✓
Skilled workers of agriculture, forestry, and fisheries		✓	✓
Skilled workers or carpenters	✓	✓	✓
Operation and installer for plant and machine		✓	✓
Basic jobs	✓	✓	✓
Agriculture, forestry, and fisheries		✓	✓
Mining and quarrying		✓	✓
Manufacturing		✓	✓
Electricity, gas, steam and air conditioning	✓	✓	✓
Water supply; sewerage, waste management, and recovery activities			✓
Construction		✓	✓
Wholesale and retail trade, repair of motor vehicles and motorcycles		✓	
Transport and storage		✓	✓
Accommodation and service activities for food and beverage		✓	
Information and communication			✓
Financial and insurance activities/takaful		✓	✓
Real estate activities		✓	✓
Professional, scientific and technical activities		✓	✓
Administrative and support services activities	✓	✓	✓
Public administration and defense; compulsory social security activities		✓	✓
Education		✓	✓
Human health and social work activities		✓	✓
Arts, entertainment and recreation		✓	
Other service activities		✓	✓
Household activity as an employer		✓	✓
Offshore bodies and organizations		✓	
Employer			✓
Employee		✓	✓
Self-employed		✓	✓
Family workers without salary	✓	✓	✓
Still schooling		✓	✓
Graduate		✓	✓
Have not attended school		✓	✓
Never go to school		✓	✓
Preschool education	✓	✓	✓
Primary education		✓	✓
Low secondary education		✓	✓
Upper secondary education		✓	✓
Pre-university	✓	✓	✓
Special and technical skill certificate program		✓	✓
First level tertiary education at the certificate/diploma level		✓	✓
First level tertiary education at the bachelor		✓	✓
No certificate	✓	✓	
Primary school evaluation test or equivalent		✓	✓
Lower secondary assessment or equivalent		✓	✓
Malaysian certificate of education or equivalent			✓
Malaysian higher school certificate or equivalent	✓	✓	✓
Certificate of special or technical skills		✓	✓
Certificate of Polytechnic /University/bodies that give recognition or equivalent		✓	✓
Diploma / Advanced Diploma in specialized or technical skills		✓	✓
Diploma in Polytechnic / University or equivalent		✓	✓
Diploma / Advanced Diploma		✓	
Building type		✓	✓
Visibility		✓	✓
Public transport		✓	✓
Parking space		✓	✓
Type of access to entrance			✓
Type of area (city or town)	✓	✓	✓

Hence, the time taken of each classifier will be further discussed. Table VIII shows the average training time taken by each classifier. It is shown that *xgbTree* and *xgbLinear* are the contributors for the long time taken by RF family. The average time taken by *xgbTree* is 26.4s and the average time taken for *xgbLinear* is 21.37s. It is good to see that the training time of the top classifier, *C5.0* in BST family is only 1.27s. The slowest classifier as shown is *glmStepAIC*, the average time taken is 26.9s.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

These three slower classifiers took more than 20s to train a model. In the third slowest family, RF, *RRF* is the only classifier that has longer training time than the others. The outcome of this study shows that the training speed of the classifiers from a family is different. The time taken is varied and not consistent even the classifiers come from the same family. The following will discuss the time complexity of the classifiers. Time complexity is the computational complexity that describes the amount of time it takes to run an algorithm. Commonly, the time complexity is expressed using the big O notation, e.g., $O(n)$, $O(logn)$, $O(n^{\alpha})$ and others, where n is the input size in units of bits needed to represent the input. A typical big O complexity graph is indicated by the time at the y-axis and the number of data input at the x-axis. Therefore, in this paper, the number of features will be used to represent the x-axis as the training data size in this work is fixed. Figure 2 shows the time complexity of each classifier. The slowest five classifiers were labeled in the figure and since most of the classifiers have short training time, their difference in time complexity is insignificant. From Figure 2, it clearly shows that there is one classifier behave much more different than the others. The differently behaved classifier is *glmStepAIC*, the classifier with the longest training time. It appears to have the pattern of $O(n^2)$, as the number of input (number of features) increase, the time it uses to run the algorithm increase dramatically. The rest classifiers seem to have a most likely constant $O(1)$ or $O(logn)$ graph pattern. However, from Table VIII, few classifiers are noticed to have a slightly longer time taken than the other classifiers. Owing to the large range of the training time, a graph is plotted, excluding the slowest *glmStepAIC* and other classifiers with less than 3 seconds of training time. The classifiers used less than 3 seconds to predict the sales performance and the time taken for training is consistent regardless of the number of inputs.

Table -VIII: Average time taken for classifiers and families

Family	Classifier	Time taken for classifier	Time taken for family
DA	lda2	0.72	2.25
	sda	0.65	
	stepLDA	5.88	
	stepQDA	4.78	
	fda	0.83	
	pda	0.65	
NNET	rbfDDA	1.14	1.77
	mlp	1.45	
	avNNet	2.85	
	mlpWeightDecay	2.89	
	nnet	1.11	
	pcaNNet	1.20	
DT	rpart	0.70	0.96
	rpart2	0.71	
	C5.0Tree	0.68	
	ctree	0.78	
	ctree2	1.03	
	J48	1.85	
RL	PART	1.34	1.65
	C5.0Rules	0.66	
	JRip	3.88	
	OneR	0.71	
BST	C5.0	1.27	16.35
	xgbTree	26.40	
	xgbLinear	21.37	
BAG	treebag	1.11	1.11
RF	rf	1.48	6.03
	RRF	17.58	
	parRF	1.55	
	RRFglobal	3.49	
GLM	bayesglm	0.80	13.86
	glmStepAIC	26.93	
NN	knn	0.61	0.61
PLSR	pls	0.62	0.62
LMR	multinom	0.73	0.73
MARS	gcvEarth	0.77	0.77
OM	pam	0.65	0.93
	gaussprRadial	1.21	

Figure 3 shows the classifiers in the experiment that have training time more than 3s. RRF shows the $O\sqrt{n}$ graph pattern. As the number of features increases, the training time increases gradually. However, stepLDA, stepQDA, RRFglobal, and JRip show the $O(nlogn)$ graph pattern. Meanwhile, xgbTree and xgbLinear show the $O(logn)$ graph pattern.

V. CONCLUSION

In conclusion, this paper presents the evaluation of 38 classifiers from 13 families over the 20 features subsets extracted from a telecommunication retail data to predict the sales performance of the business and hence determine the location of the new openings. The retail data used in this study consists of the features of location, property, population, economic, education and outlet characteristics. The response variable was the sales performance, where it had been discretized into low or high sales performance. This paper presented a classification problem that predicted the sales performance in order to make the decision for new outlet openings. The prediction indicates the potential of a location, when the prediction is low sales performance, the business owner might need to reconsider other better location; if the prediction is high sales performance for a new location, investors can start to investigate in a more detail manner to make decision for a new opening at the new location.

In the experiment result, it was shown that the top 3 feature selection methods are prediction step in VSURF, random search and fuse learner with strategy. There are 16 features selected at least once by the three feature selection methods across the 5 times experiment. The intersect features gave a hint to the decision-makers or investors that the features bring impact to the sales performance of a given location. The top 3 machine learning families are BST, RF, and BAG. All the classifiers from these families outperform others. The three families have one common characteristic, that they are ensemble method on decision trees. The ensemble method shows that the performance of several classifiers is better than individual classifiers. The top 3 classifiers are *C5.0*, *rf*, and *parRF*. The cross-over combination of the top feature selection and top classifiers yields the AUC of more than 0.75. The highest AUC is 0.8354 produced through the combination of random search-*parRF*. In addition, this work has achieved the objective of the paper, that is machine learning is a feasible approach for retail site selection besides the conventional method.

Retrieval Number: D7186118419/2019©BEIESP
DOI:10.35940/ijrte.D7186.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

In conclusion, machine learning approach is easy to implement in retail site selection. It saves costs, time and reduces the labour effort in investigation of a new potential site for business.

Application of this process in other types of businesses and investigate the different features can be performed as future work. The effect of the company size or the number of retail stores on the prediction performance can be considered and investigated in the future.

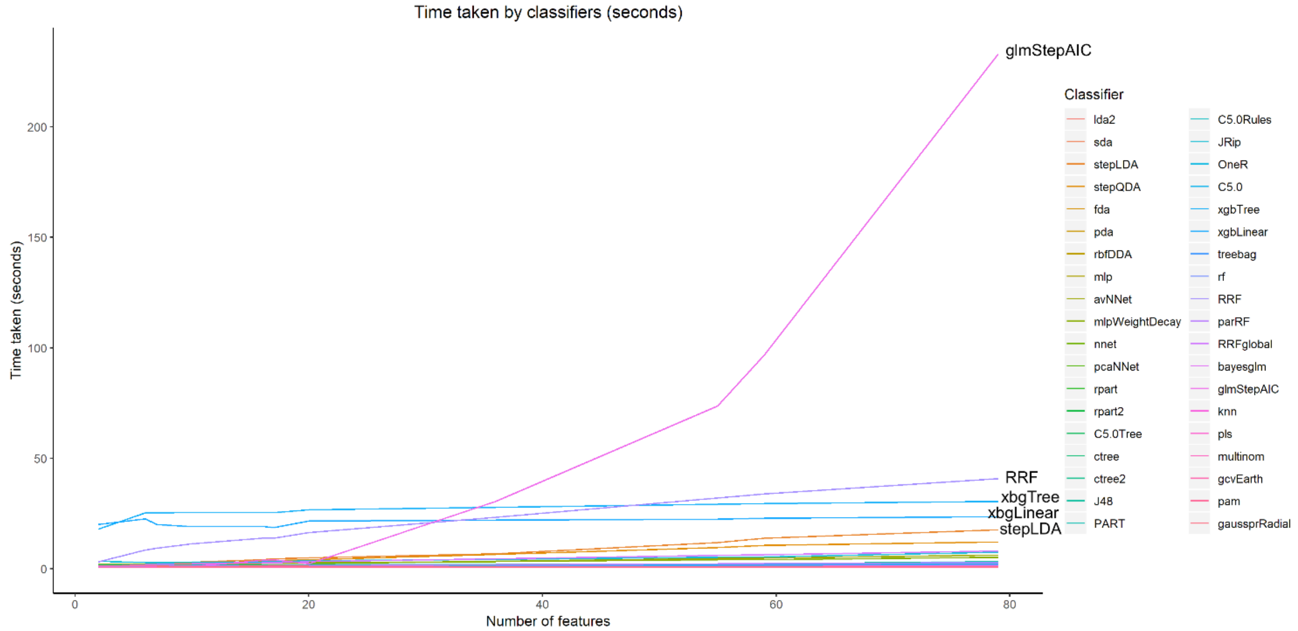


Fig. 2. Time complexity graph of classifiers

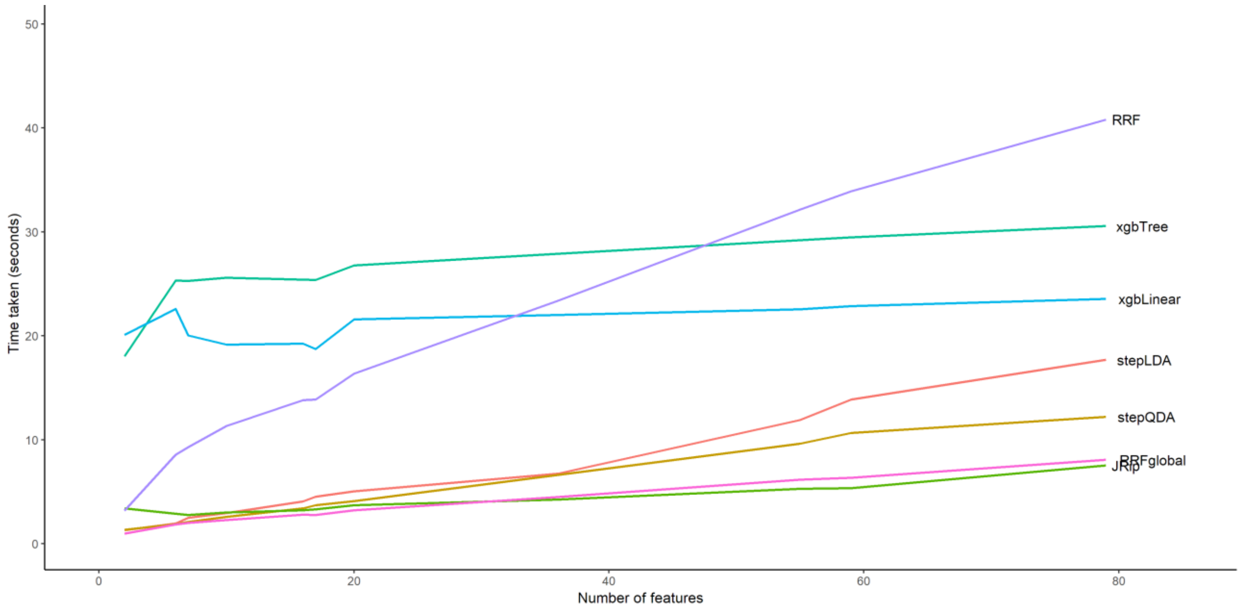


Fig. 3. Time complexity graph of classifiers that have training time more than 3s (excluding *glmStepAIC*)

ACKNOWLEDGMENT

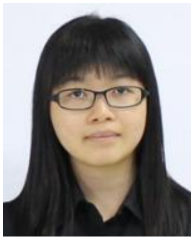
This work was supported by the LOCATIC project funded by Telekom Malaysia under Grant No MMUE/160021.

REFERENCES

1. C.-Y. Ting, C. C. Ho, H. J. Yee, and W. R. Matsah, "Geospatial Analytics in Retail Site Selection and Sales Prediction," *Big Data*, vol. 6, no. 1, 2018, pp. 42–52.
2. Esri, "GIS and Business Intelligence: The Geographic Advantage." 2016.
3. T. Turk, O. Kitapci, and I. T. Dortyol, "The Usage of Geographical Information Systems (GIS) in the Marketing Decision Making Process: A Case Study for Determining Supermarket Locations," *Procedia. Soc. Behav. Sci.*, vol. 148, 2014, pp. 227–235.
4. D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo, "Geo-spotting: mining online location-based services for optimal retail store placement.," *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 13*, Aug. 2013, pp. 793–801.
5. N. Roig-Tierno, A. Baviera-Puig, J. Buitrago-Vera, and F. Mas-Verdu, "The retail site location decision process using GIS and the analytical hierarchy process," *Appl. Geography*, vol. 40, 2013, pp. 191–198.
6. H. Erbıyık, S. Özcan, and K. Karaboğa, "Retail Store Location Selection Problem with Multiple Analytical Hierarchy Process of Decision Making an Application in Turkey," *Procedia Soc. Behav. Sci.*, vol. 58, 2012, pp. 1405–1414.
7. M. Shaheen and M. Z. Khan, "A method of data mining for selection of site for wind turbines," *Renewable and Sustainable Energy Reviews*, vol. 55, 2016, pp. 1225–1233.

8. J. Brewer, D. P. Ames, D. Solan, R. Lee, and J. Carlisle, “Using GIS analytics and social preference data to evaluate utility-scale solar power site suitability,” *Renewable Energy*, vol. 81, 2015, pp. 825–836.
9. C. Franco, M. Bojesen, J. L. Hougaard, and K. Nielsen, “A fuzzy approach to a multiple criteria and Geographical Information System for decision support on suitable locations for biogas plants,” *Appl. Energy*, vol. 140, 2015, pp. 304–315.
10. D. Latinopoulos and K. Kechagia, “A GIS-based multi-criteria evaluation for wind farm site selection. A regional scale application in Greece, 2015,” *Renewable Energy*, vol. 78, pp. 550–560.
11. A. K. Mishra, S. Deep, and A. Choudhary, “Identification of suitable sites for organic farming using AHP & GIS,” *Egypt. J. Remote Sensing and Space Science*, vol. 18, no. 2, 2015, pp. 181–193.
12. P. Chaudhary, S. K. Chhetri, K. M. Joshi, B. M. Shrestha, and P. Kayastha, “Application of an Analytic Hierarchy Process (AHP) in the GIS interface for suitable fire site selection: A case study from Kathmandu Metropolitan City, Nepal,” *Socioecon. Plann. Sci.*, vol. 53, 2016, pp. 60–71.
13. S. Kumar and V. K. Bansal, “A GIS-based methodology for safe site selection of a building in a hilly region,” *Frontiers of Architectural Research*, vol. 5, no. 1, 2016, pp. 39–51.
14. Z. Shan and Q. Zhu, “Camera location for real-time traffic state estimation in urban road network using big GPS data,” *Neurocomputing*, vol. 169, 2015, pp. 134–143.
15. J. Lin, R. Oentaryo, E.-P. Lim, C. Vu, A. Vu, and A. Kwee, “Where is the Goldmine? Finding promising business locations through facebook data analytics.,” *Proceedings of the 27th ACM Conference on Hypertext and Social Media - HT 16*, 2016.
16. K. Gdoura, M. Anane, and S. Jellali, “Geospatial and AHP-multicriteria analyses to locate and rank suitable sites for groundwater recharge with reclaimed water,” *Resour. Conserv. Recycl.*, vol. 104, 2015, pp. 19–30.
17. M. Uyan, “GIS-based solar farms site selection using analytic hierarchy process (AHP) in Karapinar region, Konya/Turkey,” *Renew. and Sustain. Energy Reviews*, vol. 28, 2013, pp. 11–17.
18. N. Yıldız and F. Tüysüz, “A hybrid multi-criteria decision making approach for strategic retail location investment: Application to Turkish food retailing,” *Socioecon. Plann. Sci.*, 2018.
19. R. Suárez-Vega, D. R. Santos-Peñate, and P. Dorta-González, “Location models and GIS tools for retail site location,” *Appl. Geography*, vol. 35, no. 1-2, 2012, pp. 12–22.
20. I. Dogan, “Analysis of facility location model using Bayesian Networks,” *Expert Syst. Appl.*, vol. 39, no. 1, 2012, pp. 1092–1104.
21. R. J. Laws and D. C. Kesler, “A Bayesian network approach for selecting translocation sites for endangered island birds,” *Biological Conservation*, vol. 155, 2012, pp. 178–185.
22. S. Guo and H. Zhao, “Optimal site selection of electric vehicle charging station by using fuzzy TOPSIS based on sustainability perspective,” *Appl. Energy*, vol. 158, 2015, pp. 390–402.
23. T. Cai, S. Wang, and Q. Xu, “Monte Carlo optimization for site selection of new chemical plants,” *J. Environ. Manag.*, vol. 163, pp. 28–38, 2015.
24. L. F. Chen and C. T. Tsai, “Data mining framework based on rough set theory to improve location selection decisions: A case study of a restaurant chain,” *Tour. Manag.*, vol. 53, 2016, pp. 197–206.
25. G. Hughes, “On the mean accuracy of statistical pattern recognizers,” *IEEE Trans. Inf. Theory*, vol. 14, no. 1, 1968, pp. 55–63.
26. V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, “Recent advances and emerging challenges of feature selection in the context of big data,” *Knowl. Based Syst.*, vol. 86, 2015, pp. 33–45.
27. B. Kumari and T. Swarnkar, “Filter versus wrapper feature subset selection in large dimensionality micro array: A review,” 2011.
28. E. R. Pacheco, *Unsupervised learning with r*. Packt Publishing Limited, 2015.
29. S. Kaushik. (2016). *Introduction to Feature Selection methods with an example (or how to select the right variables?).* Analytics Vidhya [Online]. Available: https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/.
30. M. B. Kursa and W. R. Rudnicki, “Feature selection with the Boruta package,” *J. Stat. Softw.*, vol. 36, no. 11, 2010, pp. 1–13.
31. M. Kuhn, “Caret: classification and regression training,” *Astrophysics Source Code Library*, 2015.
32. J. Hausser and K. Strimmer, “Entropy: estimation of entropy, mutual information and related quantities,” *R package version*, vol. 1, no. 1, 2014.
33. P. Romanski, L. Kotthoff, and M. L. Kotthoff, “Package ‘FSelector.’” 2018.
34. B. Bischl, M. Lang, L. Kotthoff, J. Schiffner, J. Richter, E. Studerus, G. Casalicchio, and Z. M. Jones, “mlr: Machine Learning in R.,” *J. Mach. Learn. Res.*, vol. 17, no. 1, 2016, pp. 5938–5942.
35. T. Hothorn, P. Hothorn, T., Bühlmann, S. Dudoit, A. Molinaro, and M. J. Van Der Laan, “Survival ensembles,” *Biostatistics*, vol. 7, no. 3, 2005, pp. 355–373.
36. C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, “Conditional variable importance for random forests,” *BMC Bioinformatics*, vol. 9, no. 1, 2008.
37. C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, “Bias in random forest variable importance measures: Illustrations, sources and a solution,” *BMC Bioinformatics*, vol. 8, no. 1, 2007.
38. S. L. Riza, A. Janusz, D. Ślęzak, C. Cornelis, F. Herrera, J. M. Benitez, C. Bergmeir, and S. Stawicki, “RoughSets: data analysis using rough set and fuzzy rough set theories,” *R News*, 2015.
39. R. Genuer, J. M. Poggi, and C. Tuleau-Malot, “VSURF: variable selection using random forests,” *R package version*, vol. 1, no. 3, 2016.
40. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, 2002, pp. 321–357.
41. L. Torgo and M. L. Torgo, “Package ‘DMwR’. Comprehensive R Archive Network.” 2013.
42. L. Torgo, *Data mining with R: learning with case studies*. Boca Raton, FL: Chapman & Hall/CRC, 2011.
43. M. Polena, “Performance Analysis of Credit Scoring Models on Lending Club Data.” 2017.
44. Ray. (2015) *Quick Introduction to Boosting Algorithms in Machine Learning.* [Online]. Available: https://www.analyticsvidhya.com/blog/2015/11/quick-introduction-boosting-algorithms-machine-learning/.
45. P. M. Domingos, "Why Does Bagging Work? A Bayesian Account and its Implications." In KDD, pp. 155-158. 1997.
46. M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, 2014. “Do we need hundreds of classifiers to solve real world classification problems?”. *J Mach Learn Res*, vol. 15(1), 2014, pp.3133-3181.

AUTHORS PROFILE



Hui-Jia Yee was born Teluk Intan, Perak, Malaysia in 1993. She received the B.AS. degree in applied statistics from Universiti Sains Malaysia in 2016. She is currently pursuing a Master’s degree in information technology at the Faculty of Computing and Informatics, Multimedia University, Malaysia. From 2016 to 2018, she was a research scholar with the Data Science Lab, Multimedia University. She is the author of two conference proceedings and co-author for a journal paper. Her research interest includes location intelligence, machine learning, and distance-based similarity.

Choo-Yee Ting was born in Kuching, Sarawak, Malaysia in 1974. He received his B.S in Physics and M.S in Education from Universiti Teknologi Malaysia, in 1998 and 2000, respectively. He obtained in Ph.D, degree in Information Technology from Multimedia University, Malaysia, in 2009. From 2001 to 2019, he has been attached to the Faculty of Computing and Informatics, Multimedia University. He is the author of 27 journals and 46 conference proceedings. He also handled 17 research projects. His research interests include geospatial analytics, Bayesian networks, and student modeling. Dr. Ting was one of the winners for the Malaysia Big App Challenge in 2014.

Chiung Ching Ho was born in Kuching, Sarawak, Malaysia in 1977. He received the B.S. and M.S. degrees in computer science from Universiti Putra Malaysia, Malaysia, in 2000 and 2002 respectively, and a Ph.D. degree in information technology from Multimedia University, Malaysia, in 2014. From 2005 to 2017, he was a Lecturer with the Faculty of Computing and Informatics, Multimedia University. Since 2017, he has been a Senior Lecturer with the Faculty of Computing and Informatics, Multimedia University. He is the author of three book chapters, and 17 articles, and 44 conference proceedings papers. His research interests include location intelligence, text mining, and biometrics. Dr. Ho was a winner of the Big App Challenge in 2014, and was a gold medalist in the International Invention, Innovation & Technology Exhibition in 2019.