An approach based on machine learning techniques for forecasting Vietnamese consumers' purchase behaviour

Quang Hung Do[a*] and Tran Van Trang[b]

aFaculty of Information Technology, University of Transport Technology, Hanoi, Vietnam
bFaculty of Business Administration, Thuongmai University, Hanoi, Vietnam

C H R O N I C L E	A B S T R A C T
	The main goal of this study is to investigate the classification capability of several machine learning (ML) techniques, including decision tree (DT), multilayer perceptron (MLP) network, Naïve Bayes, radial basis function (RBF) network, and support vector machine (SVM) for predicting purchase decisions. The application case is related to consumer purchase decisions of domestic goods in the context of Vietnam. Firstly, factors influencing Vietnamese consumers' purchase decision of domestic products were identified. Then, data from 240 consumers in Vietnam were collected. Different classifying models based on ML techniques were developed to analyse the sampling data after the performances of the models were evaluated and compared using confusion matrix, accuracy rate and several error indexes. The results indicate that the DT(J48) obtained the highest performance with the corrected prediction percentage of 91.6667%. The findings also show that machine-learning techniques can be used to explicitly in forecasting Vietnamese consumers' purchase behaviour.

1. Introduction

International trade is the sale and trade of goods and services across international borders. Consequently, the market has become more competitive in the recent years. Customer's attitudes toward products is concerned with all aspects of the consumer's purchase decision (Ajzen, 2015). Customer's attitudes vary greatly by country and are dependent on development status (Jacoby & Morrin, 2015). The forecasting of consumers' purchase behaviour has been getting much attention since accurate predictions assist managers and retailers in meeting customer needs and achieving profitability. It is important to understand both the theoretical and practical aspects of the factors that form the behaviour of consumer purchasing decision. It was confirmed that consumer behaviour refers first to the act of buying a certain product or service. It was stated that consumers' decision-making processes are influenced by factors such as cultural factors, social factors, personal factors, and psychological factors (Kotler et al., 2009). Consumer behaviour patterns are used to describe the relationship between these factors, including the stimulus, black box consciousness, and responses responding to consumer stimuli that affect purchasing decisions of consumers. The Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1977) showed that the intention of an individual's behaviour to be influenced by two factors, including behavioural attitude, and subjective norm. These two factors directly affect the behavioural intent and then affect the actual behaviour of an individual. Theory of

Planned Behaviour (TPB) indicated that the intention is influenced by three factors, including behavioural attitudes, subjective norms, and perceived behaviour control. Also, various studies showed that purchase behaviour of consumers is effected by the following factors: ethnocentrism, perceived price, and perceived quality of domestic products (Cham, Ng, Lim, & Cheng, 2018; Diehl, Kornish, & Lynch, 2003; Erdogan & Uzkurt, 2010; Hari & Prasad, 2014; Luque-Martínez, Ibáñez-Zapata, & del Barrio-García, 2000; Watson & Wright, 2000). The above elements and factors are included in our study to develop the research model for forecasting Vietnamese consumers' purchase behaviour.

In order to predict purchase behaviour, various approaches including statistical and multivariate analysis have been applied. Tien (2019) used binary logistic regression model to analyse the factors influencing Vietnamese consumers' purchase decision of domestic apparel. Zuo (2016) proposed an operational approach to the construction of a Bayesian network (BN) to predict purchase behaviour. The results revealed that the BN showed greater forecast accuracy than linear discriminant analysis and logistic regression analysis. Hester (HESTER, 1989) investigated the relationship between consumer attitude and consumer behaviour in an apparel purchase. Surveys were conducted with 3,766 consumers in nine locations in the eastern U.S.A. Consumers were interviewed about their attitudes toward domestic versa imported clothing and their awareness of the country of origin of their purchase. However, to develop more effective prediction models rather than traditional purchase behaviour forecasting approaches, it is better to utilize computational intelligence models (Cirqueira, Hofer, Nedbal, Helfert, & Bezbradica, n.d.). The computational intelligence approach has the capability to explore the nonlinear relationship and discover hidden knowledge from dataset. As a result, the approach has been applied to a number of practical problems in various scientific disciplines.

The purpose of the study is to develop ML-based models in dealing with customer purchase prediction in Vietnam context. The modelling is supported by a questionnaire, surveying individuals for information on ethnocentrism, perceived price and perceived quality. Other than that, this study examines the validity of several artificial intelligence (AI) techniques including decision tree (DT), multilayer perceptron (MLP) network, Naïve Bayes, radial basis function (RBF) network and support vector machine (SVM) in prediction of the purchasing behaviour of customers. The remainder of the paper is organized as follows. The brief introduction of the five ML techniques is represented in Section 2. Section 3 is devoted to the description of collected dataset. Section 4 presents results and discussions. Finally, conclusions are provided in Section 5.

2. Preliminaries

In the study, five AI techniques are used for evaluating spatial and temporal variations in purchase prediction. These techniques are briefly discussed as follows.

2.1. Multilayer perceptron (MLP) network

ANNs are the form of artificial intelligence, which is based on the function of human brain and nervous system. A neural network in which activations spread only in a forward direction from the input layer through one or more hidden layers to the output layer is known as a multilayer feed-forward network. For a given set of data, a multi-layer feed-forward network can give a good non-linear relationship. Studies have shown that a feed-forward network even with only one hidden layer can approximate any continuous function (Hornik, Stinchcombe, & White, 1989; Funahashi, 1989). Therefore, a feed-forward network is an attractive approach (NøRgaard, Poulsen, & Ravn, 2000). Figure 3 shows an example of a feed-forward network with three layers. In Figure 3, R, N, and S are the numbers of input, hidden neurons, and output, respectively; iw and hw are the input and hidden weights matrices, respectively; hb and ob are the bias vectors of the hidden and output layers, respectively; x is the input vector of the network; ho is the output vector of the hidden layer; and y is the output vector of the network. The neural network in Fig. 1 can be expressed through the following equations:

$$h_{oi} = f\left(\sum_{j=1}^{R} iw_{i,j}.x_j + hb_i\right), \; for \; j = 1,...,N \tag{1}$$

$$y_i = f\left(\sum_{k=1}^{N} hw_{i,k}.ho_k + ob_i\right), \quad for\ i = 1,...,S \tag{2}$$

where f is an activation function. When implementing a neural network, it is necessary to determine the structure in terms of number of layers and number of neurons in the layers. The larger the number of hidden layers and nodes, the more complex the network will be. A network with a structure that is more complicated than necessary over fits the training data (Caruana, Lawrence, & Giles, 2001). This means that it performs well on data included in the training set, but may perform poorly on that in a testing set.
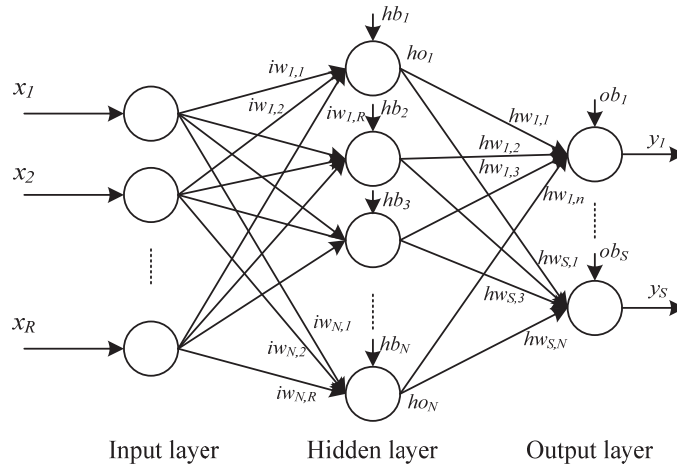

Fig. 1. A feed-forward network with three layers

Once a network has been structured for a particular application, it is ready for training. Training a network means finding a set of weights and biases that will give desired values at the network's output when presented with different patterns at its input. When network training is initiated, the iterative process of presenting the training data set to the network's input continues until a given termination condition is satisfied. This usually happens based on a criterion indicating that the current achieved solution is good enough to stop training. Some of the common termination criteria are sum of squared error (SSE) and mean squared error (MSE). Through continuous iterations, the optimal or near-optimal solution is finally achieved, which is regarded as the weights and biases of a neural network. Suppose that there are m input-target sets, $x_k - t_k$ for $k = 1,2,...,m$ for neural network training. Thus, network variables arranged as *iw, hw, hb,* and *ob* are to be changed to minimize a cost function. E, such as the MSE between network outputs, y_k, and desired targets, t_k, is as follows:

$$MSE = \frac{1}{m}\sum_{k=1}^{m} e_k^2 = \frac{1}{m}\sum_{k=1}^{m} (t_k - y_k)^2 \tag{3}$$

2.2. Radial basis function (RBF) network

RBF network is a kind of kernel function network that uses kernel functions, located in different neighbourhoods of the input space. The architecture of RBF network includes three layers: the input layer, the hidden layer, and the output layer, as shown in Fig. 2. Although the structure of Radial Basis Function (RBF) neural network is rather simple, the network has a strong generalization ability (Jiang, Cao, & Chen, 2016). The RBF neural network has shown a good classification and approximation performance in various applications (Batool et al., 2013; Guan, Zhu, & Song, 2016).
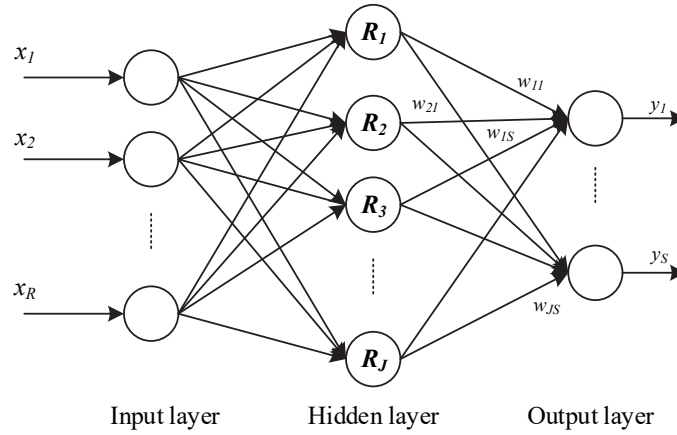
Fig. 2. A RBF network

As shown in Fig. 2, the estimated output is a weighted summation utilizing the following equation:

$$y_s = \sum_{j=1}^{J} w_{js} R_j(x), \quad s = 1,2,...S$$

where S denotes the number of outputs, J is the number of nodes in the hidden layer and w_{js} is the connection weight between j-th node of the hidden layer and s-th node of the output layer. There are several radial basis functions, the most commonly used one is as follows:

$$R_j(x) = \exp\left(-\frac{\|x - c_j\|}{2\sigma_j^2}\right), \quad j = 1,2,...J$$

where x is the input pattern vector, each input is represented by the N-dimensional vector; c_j and σ_j are the center and the width of RBF, respectively; $\|x - c_j\|$ is the norm of the vectors x and c_j, which can be considered as the distance between the vectors x and c_j. Through the RBF network, the relationship between the input and output is established. The design and training of an RBF through the estimation of three kinds of parameters, including the center and width of radial basis functions and the connection weights.

2.3. Decision Tree (DT)

Decision tree is also one of the most used intelligence techniques because of its simplicity to understand and interpret the results. A DT classifies original input variables into subgroups that construct a tree with a root node, internal nodes and leaf nodes. A decision tree (DT) can be considered as a hierarchical model composed of decision rules that recursively split independent inputs into homogenous sections (Myles, Feudale, Liu, Woody, & Brown, 2004). The aim of constructing a DT is to explore the set of decision rules that can be used to predict outcomes from a set of input variables. Applying a DT on a dataset would predict the target variable of a new dataset record. A DT is also called a regression or classification tree if the target variables are continuous or discrete, respectively (Debeljak & Džeroski, 2011). The DT can give the idea of importance of an attribute in a dataset.

2.4. Support Vector Machine (SVM)

SVM is a supervised learning method influenced by advances in statistical learning theory (Abe, 2010). SVM has been successfully applied to various applications in classification and recognition problems. Using training data, SVM maps the input space into a high dimensional feature space. In the feature space, the optimal hyper plane is identified by maximizing the margins or distances of class boundaries. The training points that are closest to the optimal hyper plane are called support vectors. When the

decision surface is obtained, it then can be used for classifying new data. Consider a training dataset of feature-label pairs (x_i, y_i) with $i = 1,...,n$. The optimum separating hyper plane is represented as:

$$g(x) = sign\left(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x_j) + b\right),$$

(4)

where $K(x_i, y_j)$ is the kernel function; α_i is a Lagrange multiplier; and b is the offset of the hyper plane from the origin. This is subject to constraints $0 \le \alpha_i \le C$ and $\sum \alpha_i y_i = 0$, where α_i is a Lagrange multiplier for each training point and C is the penalty. Only those training points lying close to the support vectors have nonzero α_i. However, in real-world problems, data are noisy and there will be no linear separation in the feature space. Hence, the optimum hyper plane can be identified as:

$$y_i(w.x_i + b) \ge 1 - \zeta_i, \zeta_i \ge 0$$

where w is the weight vector that determines the orientation of the hyper plane in the feature space; ζ_i is the i th positive slack variable that measures the amount of violation from the constraints.

2.5. Naive Bayes Classifier

A Naive Bayes classifier is based on Bayes' theorem and the probability that a given data point belongs to a particular class (Han, Kamber, & Pei, 2006). Assume that we have m training samples (x_i, y_i), where $x_i = (x_{i1}, x_{i2},..., x_{in})$ is a n-dimensional vector and y_i is the corresponding class. For a new sample x_{tst}, we wish to predict its class y_{tst} using Bayes' theorem:

$$y_{tst} = \arg\max_y P(y|x_{tst}) = \arg\max_y \frac{P(x_{tst}|y)P(y)}{P(x_{tst})}$$

(3)

However, the above equation requires estimation of distribution $P(x|y)$, which is impossible in some cases. A Naive Bayes classifier makes a strong independence assumption on this probability distribution by the following equation:

$$P(x|y) = \prod_{j=1}^{n} P(x_j|y).$$

(4)

This means that individual components of x are conditionally independent given its label y. The task of classification now proceeds by estimating n one-dimensional distributions $P(x_j|y)$.

3. Dataset

Data were gathered from 240 consumers in Vietnam in 2018. These customers represent different geographic, cultural, and commercial backgrounds. Apparel domestic products were chosen as the domestic product category in this study. The chosen product category in which a domestic alternative was available. A total of 14 items were used to measure three independent variables of the research model. Table 1 shows the three independent variable and 14 question items. Level of agreement with each of the items using 5 point Likert scale from category "strongly disagree" to category "strongly agree". Other demographic variables including age, gender, income, education were also included in the model. The other independent variables are "Willingness to pay (WTP)" and "Having child". These independent variables are included in Table 2. The dependent variable is consumers' purchasing decision (0 = if respondent is irregular buyer, 1= if respondent is regular buyer).

Table 1
Independent variables and their measured items

Independent variable	Item
Perceived price (PP)	The price of domestic products is more acceptable than foreign manufactured products (PP1).
	Compared to the quality, the price of domestic products is cheaper (PP2).
	The amount of money to buy domestic products is perfectly suited to me personally (PP3).
	The amount of money to buy domestic products compared with foreign manufactured products is reasonable (PP4).
Perceived quality (PQ)	Seam strength is not inferior to foreign manufactured products (PQ1).
	The fabric is not inferior to foreign manufactured products (PQ2).
	Brand prestige is not inferior to foreign manufactured products (PQ3).
	Production techniques are not inferior to foreign manufactured products (PQ4).
Consumers ethnocentrism (CE)	Buying foreign manufactured products are a bad behavior of Vietnamese people (CE1).
	Vietnamese had better buying goods made in Vietnam (CE2).
	Buying foreign manufactured products is contributing to the loss of some Vietnamese workers (CE3).
	Buying foreign manufactured products will help other countries get rich (CE4).
	Foreign manufactured products cause harm to the domestic firms (CE5).
	It should only buy foreign manufactured products when it cannot be produced domestically (CE6).

Table 2
Individual-level variables

Variable	Description
Age	Age of respondents in years
Gender	Dummy = 1 if the respondent is male
Educational attainment	Dummy = 1 if "High school or higher"
Income	Dummy = 1 if "USD 500 or higher"
Having child	Dummy = 1 if "Yes"
Willingness to pay	Categorical variable: 1 = "0%"; 2 = "Less than 5%"; 3 = "6-10%"; 4 = "11-15%"; 5 = "16-20%"; 6 = "more than 20%"
Consumption	Categorical variable: 1 = "None"; 2 = "Less than USD30 per month"; 3 = "USD30-50 per month"; 4 = "More than USD50 per month"

For reliability and validity of items, Cronbach's alpha values were calculated. the lowest Cronbach's alpha value is 0.878, exceeding the cut value of 0.60 recommended (Hair, Anderson, Tatham, & Black, 1998). The value of Kaiser-Mayer-Olkin (KMO) was 0.816 (in the range from 0.5 to 1.0) which means that data is significant for conducting a factor analysis. All loadings of variables are higher than the 0.50.

4. Results and Discussions

All models were coded in Python environment. In order to avoid over-fitting problem, the 10-fold cross validation was utilized. For each technique, various sets of parameters were tried to obtain the best architecture of each classifying model. To evaluate the performance of the forecasting model, several performance criteria were used. These criteria were applied to know how well the developed models worked. They are as follows: the percentage of accurate and inaccurate classification, mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), root relative squared error (RRSE) and confusion matrix. Theoretically, a forecasting model is regarded as better when these error values are smaller. The performance statistics of different techniques are represented in Tables 3 and 4. Table 3 shows the performance criteria of the developed classification model. Among developed models, DT(J48) achieved the best performance, followed by MLP, SVM, Naïve Bayes and RBF. DT(J48) obtained the smallest values in all criteria including MAE, RMSE, RAE and RRSE. According

to Table 4, DT(J48) achieved the highest correct classification rate of 91.6667 %, followed by MLP, Naïve Bayes, RBF and SVM. Other than that, DT(J48) obtained the highest performance according to all four evaluation criteria including MAE, RMSE, RAE and RRSE. Fig. 3 shows the visualization of decision tree derived from J48 method. Figure 4 represents the 2-class confusion matrixes obtained from different techniques. Actually, there are a total of 240 samples, including 119 respondents who are not irregular buyers, 121 respondents who are irregular buyers. Among investigated techniques, DT(J48) gives quite good results: 111 among 119 respondents who not being irregular buyers are correctly classified and 109 among 121 respondents who being irregular buyers are correctly classified.
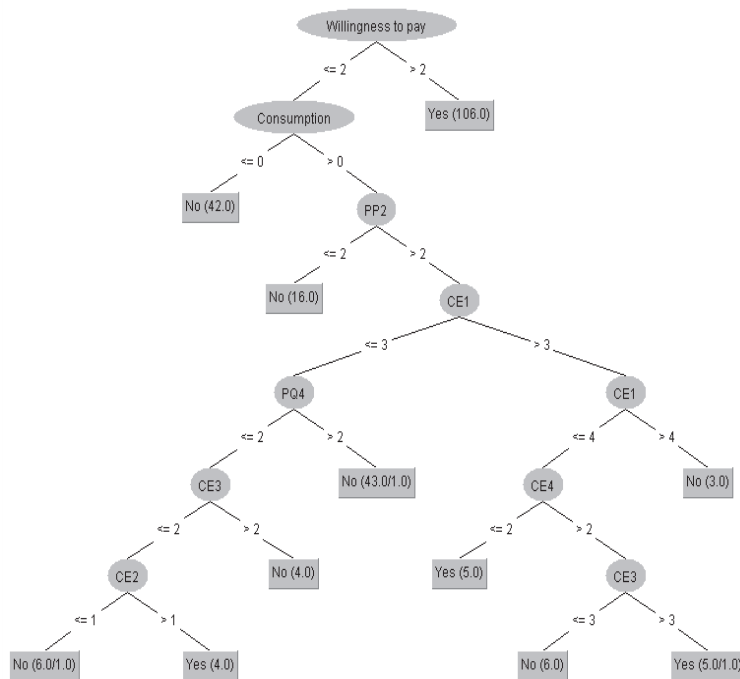
Table 3
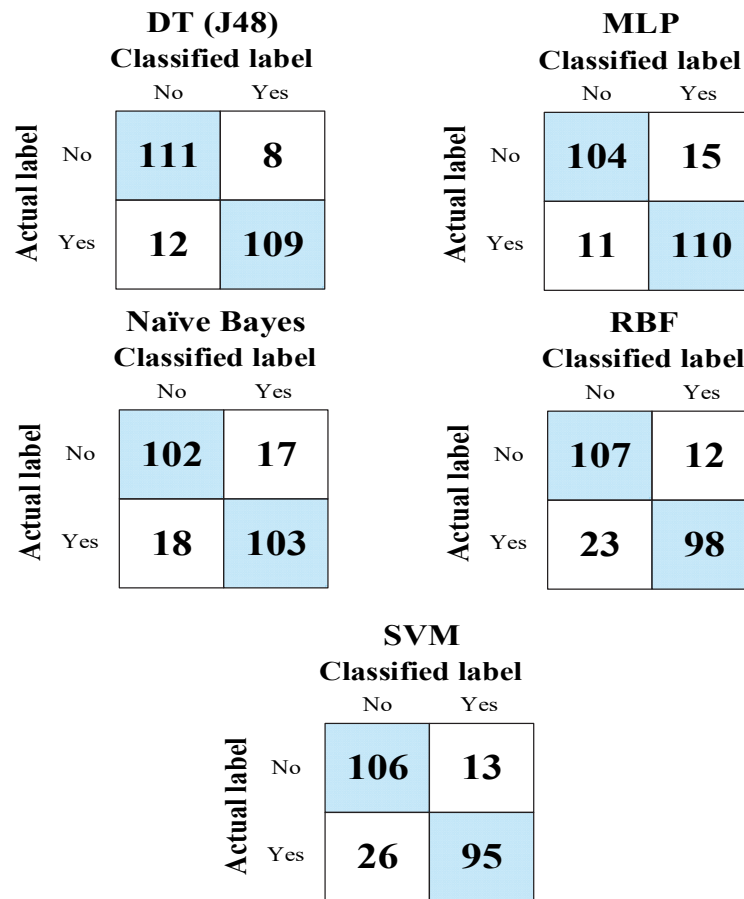Performance statistics of different ML-based models

Model	MAE	RMSE	RAE	RRSE
DT (J48)	0.0955	0.2764	19.0927%	55.2763%
MLP	0.1485	0.2839	29.7004%	56.785%
Naïve Bayes	0.1805	0.3309	36.1065%	66.1795%
RBF	0.2719	0.344	54.3725%	68.803%
SVM	0.1625	0.4031	32.5%	80.6195%

Table 4
The percentage of accurate and inaccurate classification

Model	Correctly Classified Samples	Incorrectly Classified Samples
DT (J48)	220 (91.6667%)	20 (8.3333%)
MLP	214 (89.1667%)	26 (10.8333%)
Naïve Bayes	205 (85.4167%)	35 (14.5833%)
RBF	205 (85.4167%)	35 (14.5833%)
SVM	201 (83.75%)	39 (16.25%)

Fig. 3. Visualization of decision tree

DT (J48)
Classified label

		No	Yes
Actual label	No	111	8
	Yes	12	109

MLP
Classified label

		No	Yes
Actual label	No	104	15
	Yes	11	110

Naïve Bayes
Classified label

		No	Yes
Actual label	No	102	17
	Yes	18	103

RBF
Classified label

		No	Yes
Actual label	No	107	12
	Yes	23	98

SVM
Classified label

		No	Yes
Actual label	No	106	13
	Yes	26	95

Fig. 4. Obtained confusion matrixes

5. Conclusions

Understanding consumers' purchase behaviour may provide businesses with opportunities to understand their consumer needs and improve their satisfaction. Especially, nowadays consumers demand has increased personalisation and expect products and services that cater to their individual needs and wants. Therefore, prediction of consumers' purchase behaviour as shown in this study essentially facilitate a clear view of the current state of marketing system. These findings constitute important baseline knowledge to support the development of ML-based model to predict customers' behaviour. As a methodological contribution, this paper presented a practical method to identifying and understanding the factors that influence consumers' purchase behaviour. The incorporated ML techniques include decision tree (DT), multilayer perceptron (MLP) network, Naïve Bayes, radial basis function (RBF) network and support vector machine (SVM). Stepwise, these techniques have uncovered and visualized the data structure and facilitated meaningful references with specific focuses on customers' behaviour. The findings confirmed that DT(J48) outperformed the other methods in accuracy. In this study, we provide a conceptual framework in consumer purchase prediction. There is also the possibility of profoundly exploring sequential machine learning to capture the evolving consumer behaviour, and the assessment of interpretability for purchase predictions. That would also enable the optimization of predictive models and improved human computer interaction, as customers would visualize the reasons for why they are being recommended or targeted special services. Our study is not without its limitations, one of which was the small number of explanatory variables that took part in the study. Including new variables is planned to introduce to a future study. Also, applying an optimisation approach to improve the prediction accuracy is one of the most important tasks that should be included in a future study.

References

Abe, S. (2010). Support Vector Machines for Pattern Classification. In *Abe Shigeo [eds.] Support Vector Machines for Pattern Classification - Advances in Pattern Recognition*. https://doi.org/10.1007/978-1-84996-098-4

Ajzen, I. (2015). Consumer Attitudes and Behavior. In *Handbook of Consumer Psychology*. https://doi.org/10.4324/9780203809570.ch20

Batool, F., Adeel, S., Azeem, M., Khan, A. A., Bhatti, I. A., Ghaffar, A., & Iqbal, N. (2013). Gamma radiations induced improvement in dyeing properties and colorfastness of cotton fabrics dyed with chicken gizzard leaves extracts. *Radiation Physics and Chemistry*, *89*, 33-37.

Caruana, R., Lawrence, S., & Giles, C. L. (2001). Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. In *Advances in neural information processing systems* (pp. 402-408).

Cham, T. H., Ng, C. K. Y., Lim, Y. M., & Cheng, B. L. (2018). Factors influencing clothing interest and purchase intention: a study of Generation Y consumers in Malaysia. *The International Review of Retail, Distribution and Consumer Research*, *28*(2), 174-189.

Cirqueira, D., Hofer, M., Nedbal, D., Helfert, M., & Bezbradica, M. (n.d.). *Customer Purchase Behavior Prediction in E-commerce: Current Tasks, Applications and Methodologies*.

Debeljak, M., & Džeroski, S. (2011). Decision trees in ecological modelling. In *Modelling complex ecological dynamics* (pp. 197–209). Springer.

Diehl, K., Kornish, L. J., & Lynch Jr, J. G. (2003). Smart agents: When lower search costs for quality information increase price sensitivity. *Journal of Consumer Research*, *30*(1), 56-71.

Erdogan, B. Z., & Uzkurt, C. (2010). Effects of ethnocentric tendency on consumers' perception of product attitudes for foreign and domestic products. *Cross Cultural Management*. https://doi.org/10.1108/13527601011086595

Fishbein, M., & Ajzen, I. (1977). *Belief, attitude, intention, and behavior: An introduction to theory and research*.

Guan, X., Zhu, Y., & Song, W. (2016). Application of RBF neural network improved by peak density function in intelligent color matching of wood dyeing. *Chaos, Solitons & Fractals*, *89*, 485-490.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multivariate data analysis (5th ed.). In *Upper Saddle River: Prentice-Hall*.

Han, J., Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques. 2001. *San Francisco: Morgan Kauffman*.

Hari, G., & Prasad, S. (2014). Factors Influencing Buying Behavior of a Selected Apparel Retailer's Customers. *Symbiosis Centre for Management Studies Pune Annual Research Journal of Symbiosis Centre for Management Studies Annual Research Journal of SCMS*.

Hester, S. B. (1989). Does attitude predict purchase behaviour of apparel products?. *Journal of Consumer Studies & Home Economics*, *13*(2), 161-174.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359-366.

Jacoby, J., & Morrin, M. (2015). Consumer Psychology. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*. https://doi.org/10.1016/B978-0-08-097086-8.22004-7

Jiang, J., Cao, D., & Chen, H. (2016). Boundary value problems for fractional differential equation with causal operators. *Applied Mathematics and Nonlinear Sciences*, *1*(1), 11-22.

Funahashi, K. I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, *2*(3), 183-192.

Kotler, P., Armstrong, G., Ang, S. H., Leong, S. M., Tan, C. T., & Yau, O. (2009). *Principles of marketing: A global perspective*. Prentice-Hall.

Luque-Martínez, T., Ibáñez-Zapata, J., & del Barrio-García, S. (2000). Consumer ethnocentrism measurement - An assessment of the reliability and validity of the CETSCALE in Spain. *European Journal of Marketing*. https://doi.org/10.1108/03090560010348498

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision

tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *18*(6), 275-285.

NøRgaard, M., Poulsen, N. K., & Ravn, O. (2000). New developments in state estimation for nonlinear systems. *Automatica*, *36*(11), 1627–1638.

Tien, L. D. (2019). Forecasting model of Vietnamese consumers' purchase decision of domestic apparel. *The European Journal of Applied Economics*, *16*(2), 109–121.

Watson, J. J., & Wright, K. (2000). Consumer ethnocentrism and attitudes toward domestic and foreign products. *European Journal of Marketing*. https://doi.org/10.1108/03090560010342520

Zuo, Y. (2016). Prediction of consumer purchase behaviour using Bayesian network: an operational improvement and new results based on RFID data. *International Journal of Knowledge Engineering and Soft Data Paradigms*. https://doi.org/10.1504/ijkesdp.2016.075976