Journal of Information Technology Management

Print ISSN: 2008-5893

Online ISSN: 2423-5059

Hybrid Weighted Random Forests Method for Prediction & Classification of Online Buying Customers

Umesh Kumar Lilhore

Prof., Chitkara University Institute of Engineering and Technology, Chitkara University, 140401, Punjab, India. E-mail: umesh.lilhore@chitkara.edu.in

Sarita Simaiya

*Corresponding Author, Associate Prof., Chitkara University Institute of Engineering and Technology, Chitkara University, 140401, Punjab, India. E-mail: sarita.simaiya@chitkara.edu.in

Devendra Prasad

Prof., Chitkara University Institute of Engineering and Technology, Chitkara University, 140401, Punjab, India. E-mail: devendra.prasad@chitkara.edu.in

Deepak Kumar Verma

Prof., Chhatrapati Shahu Ji Maharaj University. Kalyanpur, Kanpur Uttar Pradesh, India. E-mail: deepak30057@rediffmail.com

Abstract

Due to enchantment in network technology, the worldwide numbers of internet users are growing rapidly. Most of the internet users are using online purchasing from various sites. Due to new online shopping trends over the internet, the seller needs to predict the online customer's choice. This field is a new area of research for machine learning researchers. A random forest (RF) machine learning method is a widely used classification method. It is mainly based on an ensemble of a single decision tree. Online e-commerce websites accumulate a massive quantity of data in large dimensions. A Random Forest is an efficient filter in high-dimensional data to reliably classify consumer behaviour factors. This research article mainly proposed an extension of the Random Forest classifier named "Weighted Random Forests" (wRF), which incorporates tree-level weights to provide much more accurate trees throughout the calculation as well as an assessment of vector relevance. The weighted random forest algorithm incorporates the C4.5 method named a *"Hybrid Weighted Random Forest"* (HWRF) to forecast online consumer purchasing behaviour. The experimental results influence the quality of the proposed method in the prediction of the behaviour of online buying customers over existing methods.

Keywords: Weighted random forest, Machine learning, Classification, Prediction, Online customer

DOI:10.22059/jitm.2021.310062.2607 Manuscript Type: Research Paper

Introduction

Due to the several benefits i.e. multiple choices, discounts, and convenience of online shopping, it's getting increase worldwide. Consumers always play an essential role in the purchasing of any product. Purchasing any product depends on various factors i.e. customers' budget, interest, mindset, and requirement. In the above factors, mindset plays an important role in purchasing. Some of the products attract a customer to purchase it and some of them do not attract. So we can say the all the customers have different purchasing behaviors (Akbarabadi, M., & Hosseini, M. 2020).

It is very important to recognize the customer buying prediction using various factors i.e. when, what, where, why, and how. Many researchers are working in the area of customer behavior prediction. In many of the countries i.e. USA, China, India, and Japan approximately 12% of total sales come from various online e-commerce sites like Amazon, Ali-baba, flipchart, and Snap-deal, etc. With e-commerce becoming ever more widespread in the current marketplace, producers need to recognize which aspects make a purchase into such a website owner and be likely to draw attention to prospective consumers (Gajowniczek et al. 2020).

Researchers thought it would be useful to think that something is difficult to forecast a website page visitor's purchase decision as it can all have long-term effects, including a website for e-commerce which can great pick advertisements as well as find out determinants to strong profits (Buettner, R.et al. 2020).

This research paper presents a Hybrid Weighted Random Forest (WHRF) to predict the behaviors of online buying customers. This complete paper is divided into various sections which include an introduction, literature survey, problem identification, and proposed solution for customer prediction, implementation, result discussion, and finally conclusion.

Related Work

Machine learning techniques play a significant role in the perception of consumer data. The various e-commerce online platforms generate large amounts of data. The whole dataset includes high-dimensional data which requires great consideration. Several researchers have proposed various strategies for analyzing high-dimensional data, some of them being as follows-

De Caigny et al. 2018 presented a hybrid method based on the classification method for online customer's contents and review. The proposed method classifies the various sentimental of the user and helps in choosing an online product. The experiment results clearly showed the importance of various user opinions i.e. positive, negative, and neutral about the online product. Hu, X., et al. 2020, presented the user behavior modeling, recommendations, and purchase prediction during shopping festivals. To increase the user count over online shopping platforms an online shopping behavior analysis has been discussed. In this work, a collaborative filter method is used to predict whether an online customer will purchase a product or not.

Ayodeji, et al. 2020 presented predicting online shopping cart abandonment with machine learning approaches. This work discussed various machine learning methods to predict the e-commerce shopping cart abandoners. German online data sets were mainly used for this research which contents a total of 821,048 observations of various online customers.

Khanvilkar, G., & Vora, D. 2018, researchers presented refined weighted random forest and its application to credit card fraud detection. A random forest performs great in the classification of data, but during the classification voting, it assumes a common weight for it's all the classifiers. Bootstrap sampling and attributes selecting can't guarantee common decision making in Random forest and resulting in some of the classifiers have higher and some have lower weights. This research mainly covers a weighted random forest method.

Liu, Yaxi, et al. 2020, presented the impact of trust in consumer protection on internet shopping behavior: an empirical study using a large official dataset from the European Union. This work mainly uses a data analysis model based on the logistic regression method. The experimental analyses were performed on the European commission dataset. The experimental result clearly shows that user trust in customer protection had a limited effect on e-commerce users.

Naresh et al. 2020, presented to order or not to order: predicting customer grocery shopping behavior using multi-label classification techniques. This work mainly aims to predict the daily shopping probability of customers named "Short term shopping forecasting accuracy". The experimental results are satisfactory as well as help decision-makers.

Patil, V., & Lilhore, U. K. 2018, researchers presented an efficient credit card fraud detection model based on machine learning methods. This research work examined various machine learning classification techniques i.e. random forest, support vector machine, Naïve Bayes, gradient classifies over credit card fraud datasets. The proposed method and existing methods were evaluated using various performance measuring parameters i.e. precision, recall, accuracy, F1-score.

Rachid, et al. 2018, researchers presented predicting the helpfulness of online customer reviews: the role of title features. In this work, a hidden Markov model is used to detect associate customers, and later a random forest method is applied to detect the behaviors of alone customers. The experimental results clearly showed the quality of the proposed method in terms of accuracy over existing methods.

Rausch, et al. 2020, researchers presented a random forest approach for predicting the online buying behavior of Indian customers. This work developed a prediction model using the random forest method for the analysis of Indian e-commerce customers.

Goyal, R., & Manjhvar, A. K. 2020, presented a heuristic approach to online purchase prediction based on internet store visitor's classification using data mining methods. To obtain a simple and inexpensive initial solution to the problem, or at least to generate helpful patterns and facts in the data, this report concentrated on a heuristic approach for addressing the issue under circumstances of certain conceptual and analytical differences (Zeng, M., et al. 2019).

Machine learning methods for Classification

Following machine learning-based methods are widely used for classification:

Random Forest Method

The RF approach takes into consideration a causative relationship among some of the variables influencing the conduct of the online transaction. Throughout the investigation article (Parkhimenka, et al. 2017) paper developed an RFM as both a term of the offer which really fits many tree branch classifications similar to both the bootstrap samples and afterward integrates all tree predictions. RFM uses the variable value for finding the short-term bond for dependent variables to facilitate improved predictive efficiency.

In the past decade, and RFM has been gaining some attention. Researchers recommended RFM alongside a red logistical regression as unlike regression methods, tree-based approaches may not require a predetermined response-predictor connection. This tree-based procedure generates mainly a classifier model mostly on the response variable built through recursively dividing the data into subsets that are progressively less heterogeneous concerning potential confounders (Karthik, et al. 2018). Logistic regression modeling further shows the importance of each predictor in describing the consequence vector. These same incidence rates which are essential logistic regression statistics do not have to provide data relevant data design responsibilities as well as significance amongst these forecasting variables (Xuan S., et al., 2018).

Naïve Bayes

Bayes' Theorem suggests a method for us to predict the likelihood of something like a piece of information belonging to a particular class, considering one's background experience. Bayes' Theorem has been as follows in equation (1):

$$Prob_a(class|dataet) = (Prob_a(data|class) * Prob_a(class)) / Prob_a(dataset) \quad (1)$$

Where $Prob_a$(class|dataset) the likelihood of class (supplied by the data received)

A naive algorithm is an optimization technique including binary (two-class) through multiclass classification situations. It's named Stupid Bayes and Fool Bayes also because probability equations by each class are condensed which makes the estimates workable. Instead of trying to calculate the probability of each attribute value, target class value, they are believed to be linearly independent. That would be a very strong presumption that is almost impossible in actual evidence, i.e. attributes do not interfere. However, the approach performs remarkably well enough on data whereby this statement doesn't hold (Ghorbanian, F., & Jalali, M. 2020).

Proposed Hybrid Weighted Random Forest Method

The random forest approach has some of the most common classification techniques, based on machine learning. Online platforms produce huge quantities of data of large dimensions. RF might not be an appropriate filter throughout high-dimensional data to correctly identify consumer behavioral influences. Below we recommend an enhancement of Random Forest termed Weighted Random Forests (wRF) that further involves tree-level weights loads demonstrate extra precise trees throughout differential-importance forecasting as well as measurement. Through this work, we introduce an innovative weighted random forest algorithm especially combined through the C4.5 approach named a hybrid random forest system to forecast the actions of consumers buying individuals. We ensemble classifier C4.5 (Manjhvar, A. K. 2020) through a weighted random forest named hybrid process, and leverage treetop diversity to enhance the ensuing pattern.

Proposed Weighted Random Forest

The idea of cost-sensitive learning continues to follow in making random forests quite suitable besides learning from extremely imbalanced data. Although the RF classifier appears to ever be weighed against both the dominant elite, we are going to place a longer suspension for misclassifying the minorities.

The proposed method allocates a weight per class, including greater weight assigned to something like the minority class (i.e., higher cost of misclassification). The class weights become described in three positions throughout the RF algorithms. Class weights are an integral component of the refinement to reach optimal efficiency. This same RF precision calculation from outside the backpack can then be used to pick weights. The existing version of both the application implements another form, Weighted Random Forest (wRF).

Algorithm Hybrid Weighted Random Forest

In all this, we strongly recommend an expansion of Random Forest defined as Weighted Random Forests (wRF) that further integrates tree-level weights throughout order to include even more precise trees inside this estimation as well as vector significance evaluations. Through this whole paper, we propose a novel weighted random forest algorithm coupled with a C4.5 approach called a hybrid random forest system to forecast user behavior online.

Step 1- *Assure the sequence of instruction, the sequence of authentication as well as a set of measures.*

- Obtain a $K + 1$ variable as random from the available sample using the bootstrap form, including specific criteria.
- The efficiency from each dataset seems to be n, that is to assume as much as the initial dataset.
- K sets have been used as training samples of the $K + 1$ variable, as well as the remaining set is being used as a testing dataset.
- This total advanced population consists of approximately one-third of the sample size of the initial training package and comprises the evaluation package.

Step 2-*Start creating RF classifier.*

- Input K training samples, and then using the RF algorithm to construct a model;
- Composite classification algorithm constructed of the decision trees for categorization K.

Step 3- *Acquire the weighted sum by measuring the sub-classifier F-measure.*

- Select a validation collection, as well as characterize the growing sample as an individual classifier throughout the validation system by considering increasing decision trees in the forest.
- And after that get any validation set's True Positive, True Negative, False Positive, False Negative, precision rate %, and recall rate %.
- Next, evaluate the F-measure, which corresponds to the weight of each sub-classification.

Step 4- *To measure the model's efficiency feedback the evaluation package.*

Step 5- *Select the samples which are not marked. Classify all observations by just the random forest weighted towards F-measure. The outcome relies on a weighted vote within each sub-classifier's classification performance.*

Weighted Random Forest Working

In this method, the data consists of 0 or 1, binary dependent variables, including, N sample and predictor variables p. The conventional Random Forests (RF) approach will create an ensemble of n_tree classification trees to predict what will happen from the predictor variables, also every tree being trained on something like a specific validation set of N items, as well as a random subclass of selected features predictors being considered from each tree node. That original RF specification instead integrates tree-level effects across trees in equal proportion. We incorporate a normal RF optimization to construct the forest trees; although, for tree aggregation, we use productivity-dependent weights. Besides we find weighting class 'measures' from every branch of a tree to weigh quite significantly on better-performing trees.

Since weights are performance-based, trying to apply weights to the same dataset through which the weights have been measured will distort the estimation of error prediction. To stop such bias, we initially partition the dataset between testing and training ranges then utilize the training details to apply the normal RF algorithm, including trees having constructed on samples from n_tree bootstrap. While using individual out-of-bag (OOB), measurements of tree's predictive ability (including such tree-level estimation error) seem to be determined that can be used to measure weights, WJ, for each tree j=1...n_tree. Throughout the implementations of that same wRF, a testing data contained the proficiency level of the initial sample (Mutant individuals); therefore, for each branch, roughly environmental impacts of the complete sample had been in the bag and then used to construct the branch, and lower numerals have been out of the bag and it was used to test the tree's output and measure the tree's concentrations.

If the tree weights have been calculated within the dataset, we then use $_{n}$tree trees to obtain values for the remaining measurements in the independent test results, as well as calculate the votes (predicted categorizations) throughout the trees by applying the weighted w_j to all the votes. Assume v test, ij to be the subject-i vote throughout the autonomous check results for tree j, where$i = 1, \ldots, M2 \; = \; N/4$. This same weighted forecast was dependent on any trees for the subject I then becomes:

$$wP_i = \sum_{j=1}^{ntree} W_j \times V_{test,ij} \tag{2}$$

With the weighted forecast in (2), we can quantify weighting output metrics within the independent test collection, including such weighted random forest prediction error (PE_{wRF}) as well as AUC (AUC_{wRF}). When yi is the real subject class I then the PE_{wRF} estimation error can be determined defined as the weighted classification systems wCi:

$$wC_i = I(wP_i \geq 0.8) \tag{3}$$

$$PEW_{rf} = (1/mc^2) \times \sum |wc_i - Y_i| \tag{4}$$

We can also determine the weighted AUC while using weighted calculation wP_i (Equations 3 and 4) (AUC_{wRF}).

Choice of Weights

Weight w_j will be based primarily on a measure of predictive ability at the tree level for higher-performing trees. The weight of OOB training data is intuitively acceptable, reverse correlated with tree-level estimation errors. In OOB training results, identify the subjects I vote in tree j as V_{train}, and make O_{obej} an indication of both the subject's out-of-bag status in tree j. Under tree j, we describe node-level error estimation as in equation (5):

$$tP_j = 1/\sum M1\, B_{ij} \times \sum M1 \left|V_{train}, i_j - Y_i\right| \times B_{ij} \tag{5}$$

Pruning of individual trees in the forest

We did a grid search for different compositions of the α and p parameters to answer the question concerning the weighting effect of prediction quality: To check the effect of prediction quality of weighting parameters we conducted a grid check on different combinations of α and p parameters, we carried out:

α $\{0, 0.1, 0.2, 0{,}3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ These are the value of the specification of the first term (model stability) or second term (small error of the unknown dataset). p for the weight intensity (distribution) $\{0, 0.5, 1, 1.5, 2, 3, 3.2, 3.5, 4, 4.5, 5\}$-which is an exponential function. It would be simply an analytical study that would offer an insight into the phenomenon studied.

Phases in Proposed HWRF Model

The proposed system consists of four modules namely data preprocessing, data analysis, and hybrid model, and data prediction as per figure 1.

a) **Data learning & preprocessing mode:** This module is intended to clean the details. In some of the ways, not all people would be able to fill in the details while we receive data from consumers. Most of them are going to do so, and others are going to fill out the details in half. Nonetheless, we need to access the whole collection of data during analysis. The partly filled data will, in this case, be translated to full by substituting the incomplete one for the most likely one.

b) **Data Analysis mode:** This module has defined the feature that leads to purchasing behavior. The following hypotheses have been analyzed: gender-sensitive analysis of purchasing products; customers' use of promo codes; by combining assumptions 1 and 2, which sex prefers promotional code, in which store sales are more; in which storage more promotional codes are used in sales.

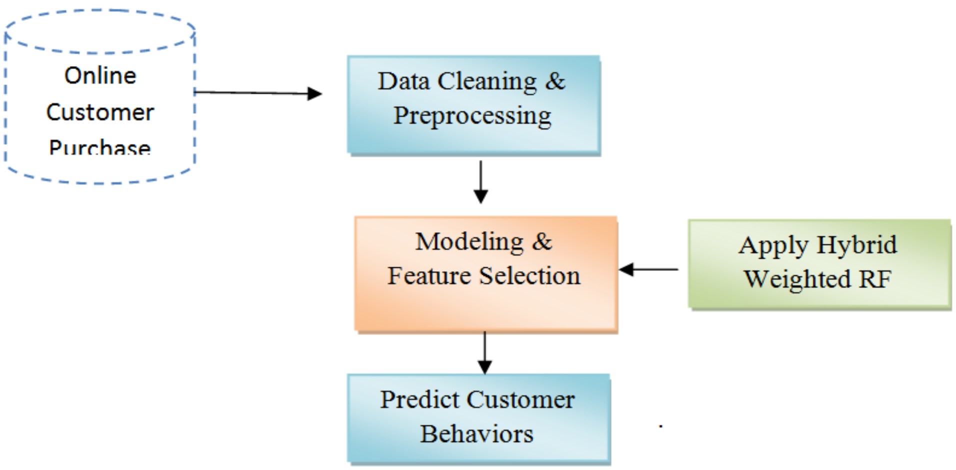

Figure 1. Phases in of Hybrid WRF

c) **Apply Hybrid Weighted Random Forest Method:** At this point, we apply the proposed hybrid weighted random forest approach to the different data sets for training and testing (Manjhvar, A. K. 2020) purposes.

- **Select the set of training:** To get K samples of the existing datasets (M) with both the size of each training set equal with that of the original training set, using a random sampling process.

- **Towards Build RF:** To generate K decision branches for a "Set of trees (Forest)," create a regression classification tree for each bootstrap training set; such trees will not be removed.

- **Concerning the growth of a tree,** this approach does not choose the best features as internal branch nodes, but rather a random collection of m is the branching procedure.

- **Establishes a strong vote:** Because the training program for each decision tree is different, random forest training will operate in tandem, thus dramatically increasing performance.
- **The RF can be formed by merging similarly qualified K decision trees:** Once the input samples are categorized, the outcomes are based on merely voting the value of each decision tree. By constructing a set of independent as well as decentralized decision trees, The RF process determines the specimens and defines a final group of the sample for each tree structure.

d) Prediction Phase throughout this research, the last subsystem is the algorithm of prediction. Within this section, we've designed to forecast which goods the consumer would most likely purchase. We're speaking about extending the forecast to random forest algorithms. Why is the algorithm preferred for random forests? The research can be performed as a collection of decision trees, and the sum of decision trees is used by random forest. It is the reason why the random forest approach is favored.

e) Evaluation Phase: After implementation, the models' performance has to be evaluated and compared. It is important to assess whether the goals, defined during the business understanding phase, are met.

Experimental Results

Data Set

The data set has been collected from Kaggle online dataset (Online buying customers). This data set mainly includes customer's age group, income, time spent on online shopping, gender, last two shopping status, customer type, etc. This data set contains 80,000 entries of various customers.

Comparison Parameters

The experiment was performed with the Proposed Hybrid Weighted Random Forest Approach and had been estimated using current Random Forest, Naïve Bayes methods, and the following parameters (Baati K, et al. 2020)

Table 1. Confusion Matrix for the proposed model

		Predicted Class	
Actual Class		Buying	Non-Buying
	Buying	TP	FN
	Non-Buying	FP	TN

- **Confusion Matrix:** A confusion m is constructed that's used to define a classification model output values on something like a set of test results on which the true values are identified.
- **Accuracy:** This is perhaps the most intuitive. It is observable by equation 6.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \tag{6}$$

- **Recall (Sensitivity)**: Recall seems to be the percentage of the right positive indicated by our system to everyone of all that is depressive. It is observable by equation 7.

$$Recall = TP/\,(TP + FN) \tag{7}$$

- **F1-score:** F1 Score needs to take both the precision as well as the recall into account. It is the precision as well as recall harmonic mean (average).

  F1 Value is better if the program has some form of compromise between accuracy (p) & recall (r). When one factor is enhanced at the cost of the other, the inverse F1 Rating is not so big. It is measurable in equation 8.

$$F1\,Score = 2 * (Recall * Precision)\,/\,(Recall + Precision) \tag{8}$$

- **Specificity-** Specificity seems to be the system appropriately considered harmful for those that are good. It is measurable in equation 9.

$$Specificity = TN/(TN + FP) \tag{9}$$

Experimental Result and Comparisons

The proposed Hybrid weighted random forest and existing Random forest method, the Naivie Bayes method was implemented using python programming and the following experimental results are calculated. A total of 80,000 customer entities are used for this experiment in which 60 % data for training and 40 % data for testing.

Confusion Matrix: The Confusion matrix is often a N x N matrix utilized for analyzing the effectiveness of the algorithm, in which N represents the size of class labels. The matrix determines the current performance measures with those expected by the learning algorithm. Confusion matrices for proposed Hybrid weighted random forest and existing Random forest method, Naivie Bayes method. It is essential to note that the methodologies are also almost comparable when the buying decision seemed to be negative.

Table 2. Confusion Matrix result for Naïve Bayes (De Caigny, et al., 2018), Random Forest(Akbarabadi, M., & Hosseini, M. 2020), and Proposed HWRF Method

Confusion MatrixNaive Bayes			Confusion Matrix Random Forest			Confusion Matrix Proposed HWRF		
a	b	Classified as	a	b	Classified as	a	b	Classified as
0.69	0.31	a= Purchase	0.68	0.32	a= Purchase	0.65	0.65	a= Purchase
0.35	0.65	b= No purchase	0.23	0.77	b= No purchase	0.19	0.81	b= No purchase

The results of figure 2 above precisely demonstrated that the developed hybrid method generates the highest rate of accuracy on test data. Compare to the conventional mechanism, the existing methods Random Forest and Naïve Bayes shows fewer results. The experimental results Table 1 confusion matrixes, Table 3, and figure 2 clearly show that the proposed HWRF method performs outstandingly in terms of accuracy over existing Random forest and Naïve Bayes methods.

Table 3. Experimental results for Naïve Bayes (De Caigny, et al., 2018), RandomForest(Akbarabadi, M., & Hosseini, M. 2020), and Proposed HWRF Method

Experimental Parameters	Experimental Methods		
	Hybrid Weighted Random Forest (Proposed)	Random forest (Existing)	Naïve Bayes (Existing)
Accuracy (%)	94.25	90.25	86.98
Sensitivity	0.1	0.23	0.34
Specificity	1.1	0.78	0.91
F1 score	0.03	0.31	0.15

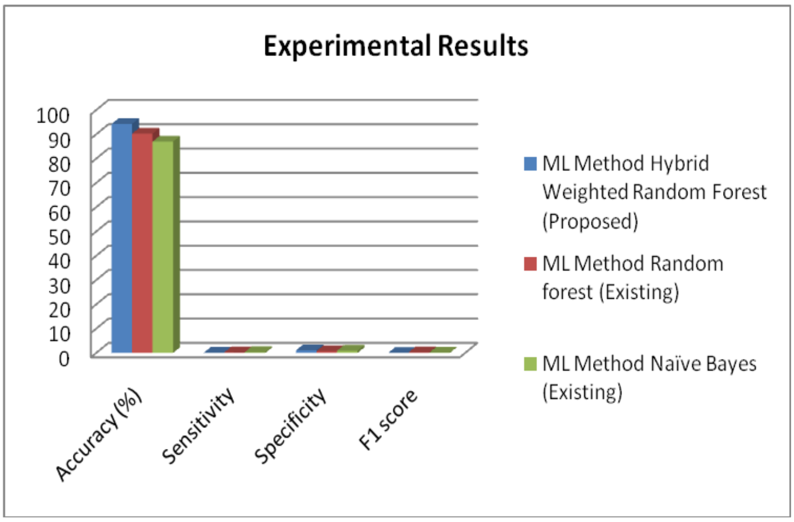

Figure 2. Experimental results Graph for Naïve Bayes (De Caigny, et al., 2018), Random Forest(Akbarabadi, M., & Hosseini, M. 2020), and Proposed HWRF Method

Conclusions & Future work

Due to the benefits and offers of e-commerce shopping more than 10 % of customers of the major population countries i.e. USA, India, and China are using online shopping. So it is always in demand for e-commerce companies to predict customer shopping behavior, but due to the dynamic nature and huge datasets, it's always challenging. This research mainly covers the hybrid weighted random forest method for online customer buying behavior prediction. Kaggle online shopping data sets are used for this research.

Random forest methods weight parameters are key-value for decision making and feature selection. Besides, using the Hybrid Weighted Random Forest model for each product segment, we tried to evaluate consumer behavioral effects on multi-channel retailers to understand whether the online shopping sector is ready for such defined product categories or if the customer prefers the conventional route. The findings reveal that the new Hybrid approach provides the maximum degree of precision on test data relative to the existing Random Forest and Naïve Bayes machine learning methods. In future research, we can test with far more criteria the efficiency of the proposed system on live real-time data and can equate it with other approaches of machine learning.

References

Akbarabadi, M., & Hosseini, M. (2020). Predicting the helpfulness of online customer reviews: The role of title features. *International Journal of Market Research*, 62(3), 272-287. https://doi.org/10.1177/1470785318819979

Gajowniczek, K., Grzegorczyk, I., Ząbkowski, T., & Bajaj, C. (2020). Weighted Random Forests to Improve Arrhythmia Classification. *Electronics*, 9(1), 99. https://doi.org/10.3390/electronics9010099

Buettner, R. (2020, August). The impact of trust in consumer protection on internet shopping behavior: An empirical study using a large official dataset from the European Union. In 2020 *IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 66-69). https://doi.org/10.1109/BigDataService49289.2020.00018

De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760-772. https://doi.org/10.1016/j.ejor.2018.02.009

Hu, X., Yang, Y., Chen, L., & Zhu, S. (2020, May). Research on a Prediction Model of Online Shopping Behavior Based on Deep Forest Algorithm. In 2020 *3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)* (pp. 137-141). https://doi.org/10.1109/ICAIBD49809.2020.9137436

Ayodeji, O. G., Kumar, V., & Kumar, S. (2020). Online retail in India: a comparative analysis of top business players. *International Journal of Indian Culture and Business Management*, 20(3), 359-384. https://doi.org/10.1504/IJICBM.2020.10023799

Khanvilkar, G., & Vora, D. (2018). Sentiment analysis for product recommendation using random forest. International Journal of Engineering & Technology, 7(3), 87- https://doi.org/89.10.14419/ijet.v7i3.3.14492

Liu, Yaxi, Dayu Cheng, Tao Pei, Hua Shu, Xianhui Ge, Ting Ma, Yunyan Du, Yang Ou, Meng Wang, and Lianming Xu. "Inferring gender and age of customers in shopping malls via indoor positioning data." *Environment and Planning B: Urban Analytics and City Science* 47, no. 9 (2020): 1672-1689. https://doi.org/10.1177%2F2399808319841910

Naresh Kumar Trivedi, Sarita Simaiya, Umesh Kumar Lilhore, Sanjeev Kumar Sharma. (2020). An Efficient Credit Card Fraud Detection Model Based on Machine Learning Methods. *International Journal of Advanced Science and Technology,* 29(05), 3414 - 3424. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/12032

Patil, V., & Lilhore, U. K. (2018). A survey on different data mining & machine learning methods for credit card fraud detection. International Journal of Scientific Research in Computer Science, *Engineering and Information Technology,* 3(5), 320-325. http://ijsrcseit.com/paper/CSEIT183565.pdf

Rachid, A. D., Abdellah, A., Belaid, B., & Rachid, L. (2018). Clustering prediction techniques in defining and predicting customer's defection: The case of e-commerce context. *International Journal of Electrical and Computer Engineering,* 8(4), 2367. http://doi.org/10.11591/ijece.v8i4.pp2367-2383

Rausch, T. M., Derra, N. D., & Wolf, L. (2020). Predicting online shopping cart abandonment with machine learning approaches. *International Journal of Market Research,* 1470785320972526. http://doi.org/10.1177/1470785320972526

Goyal, R., & Manjhvar, A. K. (2020). Review on Credit Card Fraud Detection using Data Mining Classification Techniques & Machine Learning Algorithms. *IJRAR-International Journal of Research and Analytical Reviews (IJRAR),* E-ISSN, 2348-1269.

Zeng, M., Cao, H., Chen, M., & Li, Y. (2019). User behaviour modeling, recommendations, and purchase prediction during shopping festivals. *Electronic Markets,* 29(2), 263-274.https://doi.org/10.1007/s12525-018-0311-8

Parkhimenka, U., Tatur, M., & Zhvakina, A. (2017, July). Heuristic approach to online purchase prediction based on internet store visitors classification using data mining methods. *In 2017 International Conference on Information and Digital Technologies (IDT)* (pp. 304-307). IEEE. http://doi.org/10.1109/DT.2017.8024313

Karthik, R. V., Ganapathy, S., & Kannan, A. (2018, August). A recommendation system for online purchase using feature and product ranking. *In 2018 Eleventh International Conference on Contemporary Computing (IC3)* (pp. 1-6). IEEE. https://doi.org/10.1109/IC3.2018.8530573

Xuan S., Liu G., Li Z. (2018) Refined Weighted Random Forest and Its Application to Credit Card Fraud Detection. In: Chen X., Sen A., Li W., Thai M. (eds) *Computational Data and Social Networks. CSoNet 2018. Lecture Notes in Computer Science, vol 11280. Springer, Cham*. https://doi.org/10.1007/978-3-030-04648-4_29

Ghorbanian, F., & Jalali, M. (2020, April). A Novel Hybrid Algorithm for Sentiment Analysis via Classifier Ensembles for Online Shops User Using User Generated Contents and Review. In Proceedings of the 2020 *6th International Conference on Computing and Artificial Intelligence* (pp. 165-169).https://doi.org/10.1145/3404555.3404645

Baati K., Mohsil M. (2020) Real-Time Prediction of Online Shoppers' Purchasing Intention Using Random Forest. In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020. *IFIP Advances in Information and Communication Technology*, vol 583. Springer, Cham. https://doi.org/10.1007/978-3-030-49161-1_4

Win, T. T., & Bo, K. S. (2020, November). Predicting Customer Class using Customer Lifetime Value with Random Forest Algorithm. In 2020 *International Conference on Advanced Information Technologies (ICAIT)* (pp. 236-241). IEEE.https://doi.org/10.1109/ICAIT51105.2020.9261792

Bibliographic information of this paper for citing:

Lilhore, Umesh Kumar; Simaiya, Sarita; Prasad, Devendra and Verma, Deepak Kumar (2021).Hybrid Weighted Random Forests Method for Prediction & Classification of Online Buying Customers. *Journal of Information Technology Management*, 13(2), 245-259