

Modeling online customer purchase intention behavior applying different feature engineering and classification techniques

Md. Shahriare Satu¹ · Syed Faridul Islam¹

Received: 1 August 2023 / Accepted: 2 November 2023

Published online: 14 November 2023

© The Author(s) 2023 [OPEN](#)

Abstract

In the evolution of digital technology, e-commerce sectors are gradually changing to realize customers' demands and supply required things with low cost and due time. Recently, various machine learning techniques have been used to investigate different activities of customers and estimate different characteristics and requirements of customers. The goal of this work is to propose a machine-learning model that employs multiple data analytics and machine learning techniques to manipulate customer records and predict their buying intention more precisely. In this study, we collected an online shoppers' purchasing intention dataset from a public data repository. Different feature transformation methods were employed in the primary dataset and generated its transformed datasets. Besides, we balanced the transformed datasets and detected outliers from them. Then, we applied different feature selection methods into primary and transformed-balanced datasets and again generated several feature subsets. Finally, various state-of-the-art classifiers were employed in primary, transformed, and all of their generated subsets. Then, different outcomes of the proposed model were analyzed and Random Forest was found as the stable classifier that produces more feasible results for any online shoppers' buying instances. In this work, this classifier provided the best accuracy of 92.39% and f-score of 0.924 for the Z-Score and Gain Ratio transformed subset. In addition, it gave the highest AUROC of 0.975 for the Square Root and Information Gain subset. We also found Z-Score transformation and Information Gain more reliable methods to convert online shoppers' customer intention dataset and get more feasible results from different classifiers.

Keywords Online shopping · Purchase intention · Machine learning · Feature transformation · Feature selection · Classification

1 Introduction

In the modern era, the demands of the e-commerce sector are influenced by consumers, businesses, and wholesalers day by day. Different e-commerce apps can serve individual customers from any location, any time over the Internet [5]. Thus, it shares the required information of products and services to the shopper more comfortably. In the United States, the trading of different online businesses exceeded traditional stores for the first time in 2018 [24]. The rapid usage of smart devices and social media transforming the way of shopping online. However, therefore, Google and Facebook

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s44163-023-00086-0>.

✉ Md. Shahriare Satu, shahriarsetu.mis@nstu.edu.bd | ¹Department of Management Information Systems, Noakhali Science and Technology University, Noakhali, Bangladesh.



obtained 116.3 and 55.8 billion US dollars, respectively from online advertising [11]. Therefore, these changes affected customer behaviors to make proper decisions on online purchasing.

Online shopper's purchase intention refers to the desire of the customer to buy various products or services from different e-commerce/e-business sectors. It is a crucial factor that directly affects the sales and revenue of any online retailer since there is no direct face-to-face interaction between purchasers and vendors. So, it becomes difficult to identify the personal interest of individual customers in buying any products. If any retailer is not concerned about the purchase intention of customers, they may face many problems, such as lower sales rates, lower customer purchases, high return rates, etc. In addition, they lose suitable opportunities to inspire and engage many loyal and potential customers buying and selling different products in the e-commerce sector. To understand their wishes, online retailers should engage customers and identify their requirements in different online channels. Many factors including product quality, price, availability of products, special discounts, visitor types, reliable online services, engagement on social media, etc. are important things to identify the interests of particular customers and serve them as well. However, it is one of the best ways to gather the prior information of individual shoppers and anticipate the different purchasing demands of the consumers [3]. By gaining a deep understanding of customer purchase intention, online retailers can optimize their buying experience and improve their chances of selling products.

Machine learning is useful to investigate and discover significant information from different types of data. It can be applied to the online shopper's purchasing records and extract the required factors in the e-commerce sector. In this work, we first collected online shoppers' purchase intention datasets from the University of California Irvine (UCI) machine learning repository. Different feature transformation methods like Minmax, Z-Score, and Square root methods are used to convert them into several suitable formats. Then, we employed the Synthetic Minority Over-sampling Technique (SMOTE) to balance and remove outliers using Inter Quartile Range (IQR) in these transformed datasets. Also, we implemented several feature selection methods like Correlation-Based Feature Selection (CFS) with Best First Search (BFS), Gain Ratio Attribute Evaluation (GRAE), and Info Gain Attribute Evaluation (IGAE) and generated different feature subsets from baseline and transformed datasets. Later, we applied various classifiers Naïve Bayes, Decision Tree (DT), Random Forest (RF), Simple Classification and Regression Trees (CART), Correlation-based Subspace Random Forest (CSForest), Forest-based Projection Approach (ForestPA), NBTree, Systematically Developed Forest (SysFor), Logistic Regression (LR), Logistic Model Trees (LMT), Sequential Minimal Optimization (SMO), Stochastic Gradient Descent (SGD), Bagging, and Library for Support Vector Machines (LibSVM) into the primary dataset (baseline) and other subsets to predict which classifiers are appropriate to identify the online purchase phenomena. We observed that RF is the best stable classifier to predict customer purchase intention more accurately. This proposed method helps retailers to identify significant factors of customer purchase intention more precisely.

The remaining paper is arranged as follows: Sect. 2 provides several state-of-arts of online shoppers' purchase intentions. Section 3 describes related methodologies that are used in this work. Section 4 provides the description of the online purchase dataset and describes the proposed methodology briefly. Section 5 represents the experimental results and compares this work with state-of-the-art works. Finally, Sect. 6 includes the conclusion and future plans for this work.

2 Related work

Numerous works were conducted to investigate customer intentions using machine learning and other computing methods. Gupta et al. [15] proposed a machine learning model that focused on customer segments to make decisions for predicting purchases based on the adaptive or dynamic pricing of a product. Kumar et al. [20] proposed a hybrid approach that predicted online consumer repurchase intentions by determining consumer characteristics and shopping mall attributes (with < 0.1 threshold value) using Artificial Bee Colony (ABC) algorithm. On testing the data set, AdaBoost outperformed other classification models, with 0.950 sensitivity and 97.58% accuracy, respectively. Eshak et al. [12] used machine learning and lexicon-based approaches to analyze communications in order to determine customer intention to purchase in the social-commerce sector. Sarkar et al. [26] analyzed online shoppers' behaviors by using MultiLayer Perceptron (MLP), Long Short-Term Memory networks (LSTM), and Recurrent Neural Networks (RNN) and predicted the visitor's shopping intent and website abandonment likelihood. Combining clickstream data with session information-based features improves the success rate of the system. Zheng et al. [35] represented a decision support system that categorized online browsing activities into purchase-oriented and general sessions as well as used extreme boosting machines (ELM) and browsing content entropy features to achieve 41.81% recall and an F-score of 34.35%, respectively. It implemented real-time bidding algorithms for online advertising strategies, improving the effectiveness

of advertisements and increasing last-touch attributions for campaign performance. Kabir et al. [19] analyzed online shoppers' empirical data where gradient boosting with RF gave the highest accuracy of 90.34% to predict customers' purchase intention. Shi [31] investigated different factors of online shopper's purchasing intention and variables such as time spent and page values were positively correlated as well as bounce rate and exit rate were negatively correlated with shopping intention. In that case, RF showed 89.50% training accuracy and 87.50% testing accuracy in predicting Online Shopper's Purchasing Intention. Liu et al. [23] proposed a Time-Preference Gate Network (TP-GN) where a pair of preference gates and time interval gates were added to the LSTM model to predict online users' purchase intentions. Esmeli et al. [13] dynamically used session features, start-of-art models, and scoring methods to predict the probability of early purchase intention. In this work, individual models provided good performance early purchase intention in terms of Area Under Curve (AUC) score. Gomez et al. [14] provided a customer embedding representation based on the customer's click-events during browsing sessions. This representation with an LSTM predictor outperforms the state-of-the-art approach on the three datasets. Sang et al. [27] used various sessions and visitor data to analyze the shopper's intent which showed 86.78% accuracy in that system. Bhagat et al. [8] explored the factors of consumers' online purchase intention and used a technology-based model to analyze these factors and get positive influences of artificial intelligence in consumers' buying behavior.

Several works were conducted where different data preprocessing techniques and machine learning models were used to investigate online shoppers' browsing activities/sessions/records, determine factors, and predict their purchase intentions more precisely. However, they did not properly employ more data cleaning, and feature engineering techniques as well as implement different types of classifiers in the primary dataset and its subsets. Therefore, various possible outcomes were not analyzed to bring out the most reliable machine learning pipeline for predicting online shoppers' purchase intentions.

3 Methods for predicting online customer purchase intention

Various techniques are used to create an online customer purchase intention model which are briefly described as follows:

3.1 Data transformation

Data transformation converts instances from one format to another format and makes it more feasible for different types of analysis [2]. In this work, we used several data transformation methods such as Min-Max, Z-Score, and Square root methods which are described as follows:

- *Min-max transformation* [1, 2] is scaled numeric instances between 0 and 1 to normalize data. It determines X_{new} using the following formula:

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X , X_{min} , and X_{max} are considered as the real instance, the minimum value, and the maximum value respectively. It provides equal importance to all features to predict more accurate results by machine learning models.

- *Z-score transformation* [1, 2] converts real data into another format that has 0 average and 1 standard deviation. The formula of the z-score (Z) is given as follows:

$$Z = \frac{(X - \mu)}{\sigma} \quad (2)$$

where X , μ , and σ are considered as the original data, mean, and standard deviation of the feature, respectively. It can be used for different features and outliers in the data.

- *Square root transformation* increases the value of each record with the power of two and converts its squared value by employing root operation. The formula for the square root transformation is as follows:

$$X_{new} = \sqrt{X^2} \quad (3)$$

where X , and X_{new} , are indicated as the original and transformed instances, respectively.

3.2 Data balancing

Many techniques are used to balance instances of underrepresented classes compared to the other classes. SMOTE is widely used for balancing different undersampling classes [28, 30]. This method randomly selects one of the minor classes and creates synthetic samples. Thus, it repeats until all minority classes are balanced with the major class. This method provides a linear interpolation between the selected samples and one of its k-nearest neighbors. It improves the performance of classifiers to predict different classes more accurately.

3.3 Outlier detection

Outlier detection is used to identify anomalous samples that do not properly fit with the normal/statistical distribution of a dataset [30]. In this work, we used IQR to detect outliers from the online shoppers' dataset. It defines the differences between the first and third quartiles, respectively. To manipulate IQR, the dataset is split into the first quartile (Q_1), second quartile/median (Q_2), and third quartile (Q_3). The instances which fall below $Q_1 - 1.5 \times IQR$ or above $Q_3 + 1.5 \times IQR$ are considered as outliers.

3.4 Feature selection methods

Feature selection methods [29] identify individual feature subsets and produce the expected outcomes by reducing irrelevant features. In this work, we used several feature selection techniques which are described briefly as follows:

- Correlation-based Feature Selection (CFS) [16] detects features by computing their correlation coefficients and the response variable. If any features are highly correlated with each other, they are considered as redundant features. Best-First Search (BFS) [4] is used to investigate correlations and identify the most informative features as well as provide more accurate outcomes. First, it removes a subset of features having the highest correlation with the target variable. Then, it evaluates the remaining features and provides the increment values using a heuristic function. This process is repeated until the desired number of features is selected.
- Gain Ratio Attribute Evaluation (GRAE) [16] determines the information gain from each feature to measure how well individual features separate from different classes. The information gain is divided by intrinsic information to calculate the entropy of the class distribution. This normalized information gain is called the gain ratio which ranks individual features (i.e., using the ranker search approach) in the dataset. The ranker search method uses various criteria to assign scores to the individual features and ranks them in descending order. GRAE can be used within the wrapper and filter approaches.
- Information Gain Attribute Evaluation (IGAE) [16] manipulates information gain (i.e., also called entropy) for each feature and response variable. In this case, different features having more information gain values are selected, and lower-scoring features can be removed. IGAE uses the ranker approach to identify quickly the most important features.

3.5 Machine learning classifiers

Several state-of-art classifiers [1, 29, 33] namely, NB, DT, RF, Simple CART, CSForest, ForestPA, NBTree, SysFor, LR, LMT, SMO, SGD, Bagging, and LibSVM were used in this work. These models are described briefly as follows:

- Naïve Bayes (NB) [16, 33] is worked based on Bayes theorem where individual features are considered "naïve" or independent from each other. In this case, the prior probability of features is constant. It determines the posterior probability of each class using the likelihood of the sample's features considering the given class and the prior probability of each class. It is an efficient method to solve both binary and multi-class classification problems.
- Decision Tree (DT) [6, 33] is used to analyze records and split the input space hierarchically until it reaches in a category. It contains three types of nodes namely, root, internal, and leaf, respectively. The root node has zero or more outgoing edges.

On the other hand, the leaf nodes provides the incoming edges, however no outgoing edges. Internal nodes represents two or more outgoing edges with only one incoming edge. The root and internal nodes both investigate instances based on existing features and splitting rules. It assess different new records by sorting them from root to leaf node.

- Random Forest (RF) [1, 9, 29, 33] is a bagging method where multiple DTs are combined to reduce overfitting and improve the performance of classification. In this method, each DT is trained on random feature subsets, and the predictions of all trees are combined using the voting technique to produce the final prediction.
- Classification and Regression Trees (CART) [6, 7] recursively partitions the input space into small regions until they contain different instances of a particular class. It represents a binary tree structure where internal node provides a decision of a feature and leaf node defines a class label for an instance.
- Correlation-based Subspace Random Forest (CSForest) [32] combines the functionality of RF and CFS to improve the performance of RF. First, it employs CFS into the dataset and identifies several feature subsets. Then, RF is trained using these selected features and predicts different categories more accurately.
- Forest-based Projection Approach (ForestPA) [1] is an ensemble of DTs that can manage a large amount of data and handle missing or noisy data to make robust predictions. In this model, each DT is trained with random feature subsets and split data at each node. Then, this model is used to predict different testing samples more precisely.
- NBTree [10] combines NB with DTs and is more efficient in managing continuous, nominal attributes, and missing values. It first trains NB with data and builds a DT using the most discriminatory feature. This process is repeated until the stopping criterion is met.
- Systematically Developed Forest (SysFor) [1] is a collection of multiple DTs that utilizes two voting techniques and creates a forest for predicting unlabeled records. It is not only focused on extracting patterns from high-dimensional data but also capable of handling low-dimensional data. In different datasets, it determines only those attributes that have high classification capabilities.
- Logistic Regression (LR) [2] estimates the probability of occurrence of the dependent variable for given independent variables. It uses a logistic function to predict binary outcomes by employing a linear combination of input variables. Thus, it shows an S-shaped curve to map any real values between 0 and 1. The coefficients of this combination are used to represent the maximum likelihood estimation.
- Logistic Model Trees (LMT) [21] combines with DT and LR and improves the performance classification. It constructs DT from training data and applies LR to each leaf node. In this method, LR chooses the path DT of the highest probability and makes the final decisions. It performs well on the dataset of many features and handles overfitting like other DTs.
- Sequential Minimal Optimization (SMO) [29] is an optimization algorithm in the context of Support Vector Machines (SVMs) that is fast, memory efficient, and manages different large-scale classification problems more appropriately. It splits a large quadratic problem into a series of sub-problems and analyzes them to get proper solutions. It identifies and optimizes two alpha values at a time when other alpha values are fixed. This process usually continues in small iterations until it reaches convergence.
- Stochastic Gradient Descent (SGD) [17] is useful for investigating a large number of instances and minimizing the cost of the model. This technique is faster than batch gradient descent. Thus, the weights of SGD are updated using the gradient loss function in an iterative way. In this case, it randomly selects training samples when some noises are optimized. However, SGD allows to escape local optima by finding a better global minimum.
- Bagging (Bootstrap Aggregation) [6] is an ensemble technique where a dataset is randomly sampled with replacement and creates multiple bootstrap samples. Then, a classifier (i.e. DT) is trained with each sample, and individual models are combined to predict a new sample. It decreases overfitting by reducing variance and improves the performance of a model.
- LibSVM [22] is an SVM Library that contains different tools for classification and regression. It provides different types of kernels such as linear, polynomial, radial basis function (RBF) sigmoid where any of them is employed on data based on the problem at hand. After optimizing different parameters, LibSVM trains and classifies new instances. Then, it estimates the probability of the class label by maintaining the Classification threshold.

4 Proposed methodology

The proposed machine learning model detects customer online purchase intention depicted in Fig. 1. The working steps of this model are given briefly as follows:

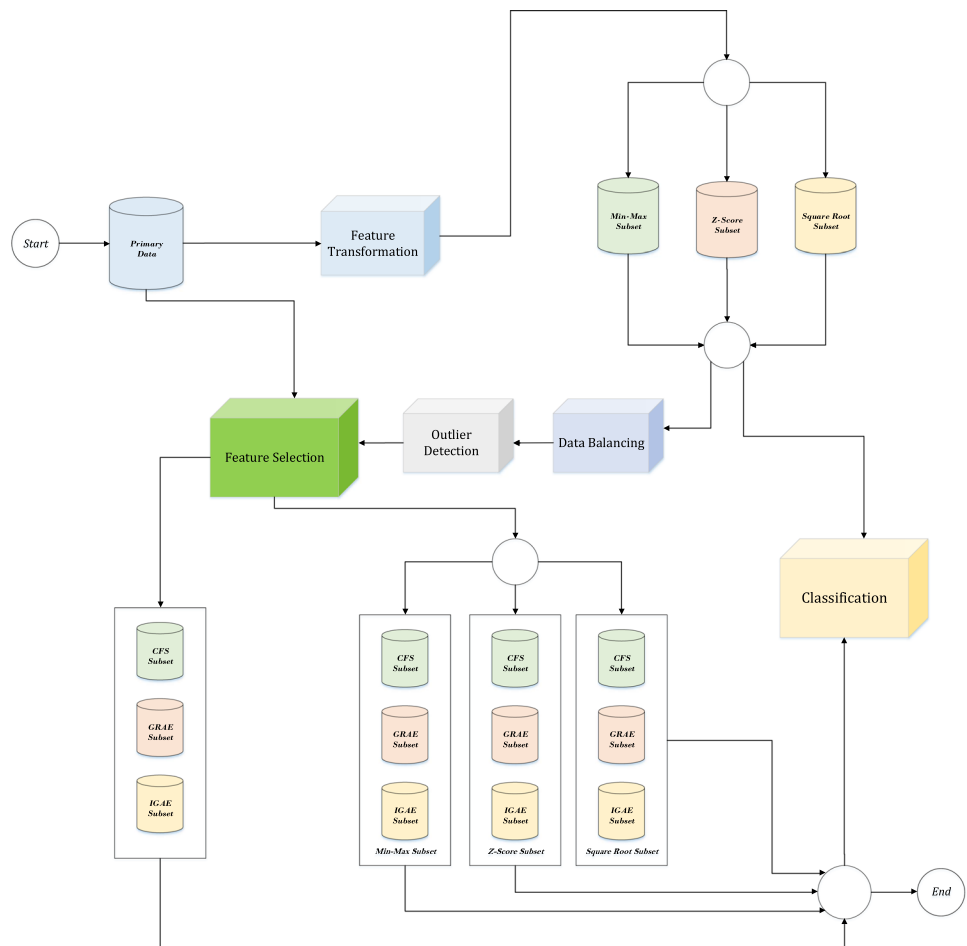
- First, we check the duplicate and missing values in the collecting dataset. If we find these things, we usually remove them from the dataset. In addition, we impute missing/wrong values with mean values of individual features or find out the appropriate optimization process for the imputation process.
- Then, several feature transformation methods such as min-max, z-score, and square root methods are used to convert the primary dataset into different formats and preserve them as transformed datasets. This process is useful for making classification more convenient ways.
- If these transformed datasets are found imbalanced, the proposed method uses a widely used SMOTE technique to balance them. In addition, the outliers are checked within this dataset using the IQR method. If it gets outliers, it may be removed or replaced with related values. Further, we also apply several feature selection methods such as CFS, GRAE, and IGAE into primary and feature-transformed datasets and generate different feature subsets from them.
- Afterward, different classifiers such as NB, RF, DT, Simple CART, CSForest, ForestPA, NBTree, SysFor, LR, LMT, SMO, SMO, Bagging, and LibSVM are employed in baseline, feature transformed datasets, and their feature subsets respectively. Several metrics including accuracy, precision, recall, F-Score, and AUROC are used to evaluate the performance of individual classifiers in these datasets.

4.1 Performance metrics

The capability of the individual classifiers for detecting customer online purchase intention was determined using evaluation metrics such as accuracy, precision, recall, f-score, and AUROC [1, 2, 6]. These metrics are computed using a confusion matrix which is a matrix-like representation of the predicted class against the actual class. Therefore, some estimated values are provided as follows:

- True Positive (TP): It estimates the positive instances of the predicted class where the actual class was also positive.

Fig. 1 Workflow of proposed methodology



- True Negative (TN): It estimates the negative instances of the predicted class where the actual class was also negative.
- False Positive (FP): It estimates the positive instances of the predicted class where the actual class was negative.
- False Negative (FN): It estimates the negative instances of the predicted class where the actual class was positive.

Then, different evaluation metrics are manipulated which are given as follows:

4.1.1 Accuracy

Accuracy is used to assess the performance of any classifier based on correctly predicted versus overall instances which are calculated using Eq. 4.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (4)$$

When the class is unbalanced, the highest accuracy is not enough to declare a classifier as the best model.

4.1.2 Precision

It calculates the ratio between true positive values and all positive predictions which is represented at Eq. 5.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

4.1.3 Recall

It computes the ratio between true positive values and all positive values of any predictive model which is represented in Eq. 6.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

4.1.4 F1-Score

F1-Score/F-Score/F-Measure is a harmonic mean of precision and recall where the value ranges from 0 to 1. The value is defined in Eq. 7 as follows:

$$\text{F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (7)$$

4.1.5 Area under the receiver operating characteristics curve (AUROC)

AUROC is an evaluation metric that constructs a result by manipulating false positive and true positive rates respectively. This value is nearest to 1 which is considered by a good model. The formula of the AUROC is given as follows.

$$\begin{aligned} \text{AUROC} &= \int_0^1 \text{TPR}(\text{FPR}) d\text{FPR} \\ &= \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx \end{aligned} \quad (8)$$

Where *TPR* is indicated as a true positive rate and *FPR* is denoted as a false positive rate.

5 Experimental result

In this work, the proposed methodology was applied to the online shoppers' purchase intention dataset and generated results. Therefore, the environmental setup, result analysis, and further discussion are described briefly as follows.

5.1 Data collection

In this study, we gathered Online Shoppers' Purchasing Intention instances from the UCI Machine Learning Repository. This dataset is generated by Sarkar et al. [26]. They collected visitors'/shoppers' records from the various e-commerce websites of leading brands including outerwear, sportswear, durable shoes, and accessories in Turkey [25]. The dataset consists of different feature vectors with 12,330 sessions so that individual users of each session for one year avoid any tendency to a specific campaign, day, user profile, or period. There are considered 1 to 10 numerical and 11 to 18 categorical features where the 'Revenue' is indicated as the class label. In these sessions, 84.5% (10,422) are negative samples that did not terminate with shopping, and the rest of them (1908) are considered positive samples that ended with shopping. Table 1 is represented these attributes with a brief explanation as follows.

5.2 Experimental setup

K-fold cross-validation is a resampling procedure that assesses machine learning models for limited data. The general hold-out method normally splits training and testing data by maintaining a ratio. However, it evaluates any model once and provides almost uncertain results. On the other hand, k-fold cross-validation creates k number of groups, and re-trained machine learning methods with different training sets in the same design. In this experiment, we investigated the online customer purchase intention dataset employing different state-of-the-art classifiers with 10-fold cross-validation (i.e., $k = 10$) in Weka version 3.8.1. The performance of every classifier was evaluated with five evaluation metrics namely accuracy, precision, recall, F-Score, and AUROC. This work was implemented on a core-i3, 11 Generation personal computer with 4 GB RAM, and Windows 11 operating system, respectively.

Table 1 Data description

S/N.	Factor	Type	Description
1	Administrative	Numeric	Number of pages of the visitor about account management
2	Administrative duration	Numeric	The total amount of time (in seconds) of the visitor on account pages
3	Informational	Numeric	Number of pages to the visitor about the website, communication and address information about the shopping site
4	Informational duration	Numeric	The total amount of time (in seconds) of the visitor informational pages
5	Product related	Numeric	Number of pages of the visitor about product-related pages
6	Product related duration	Numeric	The total amount of time (in seconds) of the visitor on product-related pages
7	Bounce rate	Numeric	Average bounce rate value of the pages visited by the visitor
8	Exit rate	Numeric	Average exit rate value of the pages visited by the visitor
9	Page value	Numeric	Average page value of the pages visited by the visitor
10	Special day	Numeric	The closeness of the site visiting time to a special day
11	Month	Category	Month value of the visit date
12	Operating systems	Category	The operating system of the visitor
13	Browser	Category	The browser of the visitor
14	Region	Category	The geographic region of the visitor
15	Traffic type	Category	Traffic source of the visitor
16	Visitor type	Category	Visitor type as "New Visitor", "Returning Visitor", and "Other"
17	Weekend	Category	The date of the visit is weekend
18	Revenue	Class	Indicating whether the visit has been finalized with a transaction

5.3 Result analysis

The overall results of different classification approaches to detect online shoppers' purchase intentions are given in the supplementary section. In addition, the maximum results (i.e., accuracy, f-score, and AUROC) for the baseline and different transformed feature subsets are given in Table 2. Now, the details analysis of these outcomes is given as follows.

When we investigated the results of individual classifiers for the balanced CFS subset (see Supplementary Table 1), LMT showed the best accuracy at 89.11% as well and SysFor provided the second-highest performance in this experiment. However, RF gave lower results than other classifiers. Besides, LMT showed the best f-score of 0.889 and CSForest represented the highest AUROC of 0.887. Then, we analyzed the performance of classifiers for the balanced GRAE subset (see Supplementary Table 3) and Simple CART provided the utmost accuracy of 89.54% and RF produced the second-highest score of 89.51% among all classifiers. From the perspective of F-Score, LMT performed the best score of 0.891 whereas RF and Bagging showed almost similar performance which is 0.890. In addition, Bagging gained the highest AUROC for this subset. Besides, the performance of classifiers for the balanced IGAE subset is shown in Supplementary Table 2. In this case, Simple CART provided the highest 90.19% accuracy and SysFor performed almost similar results of 90.16% accuracy to Simple CART among other classifiers. However, SysFor provided the best score 0.899 and Simple CART showed almost the same result (i.e., 0.898) from the perspective of F-Score. Then, LMT obtained the best AUROC score of 0.927 for this IGAE subset.

Afterward, we analyzed the outcomes of classifiers for the Min-Max Transformed CFS subset (see Supplementary Table 4), RF showed the utmost 91.52% accuracy and ForestPA almost performed similarly with 91.51% accuracy. From the view of F-measure, both RF and ForestPA provided the best result with 0.915 score whereas NBTree showed approximately the same performance (i.e., 0.914). In this case, ForestPA showed the best 0.967 AUROC among other classifiers. Again, the results of classifiers are shown for the Min-Max Transformed GRAE subset (see Supplementary Table 5). In this situation, RF performed the maximum accuracy of 91.78% and ForestPA provided the second accuracy of 91.37% in this analysis. From the point of view of F-measure, RF represented the highest outcome of 0.918 while ForestPA provided the second highest score of 0.914. Also, RF provided the maximum AUROC of 0.969 for this subset. Besides, the findings of classifiers are represented for the Min-Max Transformed IGAE subset (see Supplementary Table 6). RF showed the largest 92.16% accuracy and ForestPA provided the second highest 91.58% accuracy in this work. From the perspective of F-measure, RF also performed the best score of 0.922, and ForestPA provided the second-highest score of 0.916. RF also obtained 0.970 AUROC for the IGAE dataset.

In this circumstance, we analyzed the performance of classifiers for Z-Score Transformed CFS (see Supplementary Table 7) where RF provided the highest 91.98% accuracy and ForestPA represented the second best 91.51% accuracy in this analysis. From the view of F-Measure, RF also showed a maximum accuracy of 0.920 whereas ForestPA and Bagging represented 0.915 accuracy than other classifiers. In this circumstance, RF gave the maximum AUROC of 0.970 for this subset. Further, the findings of individual classifiers for the Z-Score Transformed GRAE dataset are provided in Supplementary Table 8. In this case, RF represented the best 92.39% accuracy and ForestPA obtained the second highest

Table 2 Maximum outcomes of different classifiers for different transformed feature subsets

Dataset type	Subset	Classifier	Accuracy (%)	Classifier	F-Score	Classifier	AUROC
Baseline	CFS	LMT	89.11	LMT	0.889	CSForest	0.887
	GRAE	Simple CART	89.54	LMT	0.891	Bagging	0.907
	IGAE	Simple CART	90.19	SysFor	0.899	LMT	0.927
Min-Max	CFS	RF	91.52	RF	0.915	ForestPA	0.967
	GRAE	RF	91.78	RF	0.918	RF	0.969
	IGAE	RF	92.16	RF	0.922	RF	0.970
Z-Score	CFS	RF	91.98	RF	0.920	RF	0.970
	GRAE	RF	92.39	RF	0.924	RF	0.972
	IGAE	RF	92.26	RF	0.923	RF	0.972
Square root	CFS	RF	92.09	RF	0.921	RF	0.973
	GRAE	RF	92.24	RF	0.922	RF	0.974
	IGAE	RF	92.22	RF	0.922	RF	0.975

Bold values indicate the highest accuracy, F-Score, and AUROC of different classifiers for individual datasets

91.41% accuracy. From the view of the F-measure, RF also gained the best score 0.924 and ForestPA obtained 0.914 score in this work. When the performance of classifiers for the Z-Score Transformed IGAE dataset (see Supplementary Table 9) is investigated, we found that RF showed the best 92.26% accuracy and ForestPA gave the second-highest accuracy of 91.46%. From the perspective of F-Score, RF also obtained the best score 0.923, and Bagging provided the second highest score 0.915. Again, RF gave the best AUROC of 0.922 scores for both GRAE and IGAE subsets.

When we analyzed the performance of classifiers for the Square Root Transformed CFS dataset (see Supplementary Table 10), RF represented the utmost 92.09% accuracy and Bagging gave the second best accuracy of 91.60%. From the view of F-Measure, RF also showed the topmost 0.921 and Bagging represented 0.916 in this analysis. Besides, RF obtained the best AUROC score of 0.973 for this subset. The performances of classifiers for the Square Root Transformed GRAE are also elaborated dataset in Supplementary Table S11, RF showed the highest accuracy of 92.24% and Bagging provided the second maximum accuracy of 91.85%. From the perspective of the F-Score, RF represented the topmost score of 0.922 and Bagging provided the second-best score of 0.919. Again, RF gave 0.974 AUROC score for this subset. The outcomes of classifiers for the Squared Root transformed IGAE dataset are given in Supplementary Table 12. In this case, RF provided the highest 92.22% accuracy and NBTree showed the second maximum 91.81% accuracy. From the view of F-Measure, RF represented the best score 0.922, and the second utmost 0.918 scores. Further, this classifier obtained an AUROC score of 0.975 for this subset.

In Table 2, Simple CART provided the best accuracy where SysFor and LMT gave the highest f-score and AUROC, for the baseline-IGAE subset respectively. In the Min-Max Transformed dataset, RF obtained the best accuracy, f-score, and AUROC for its IGAE subset. Again, RF presented the maximum accuracy of 92.39% and f-score of 0.924 for the Z-Score Transformed GRAE subset, respectively. Besides, this classifier provided the best similar AUROC score for both Z-Score Transformed GRAE, and IGAE subsets respectively. Further, RF provided the highest accuracy and AUROC of 0.975 for the Square Root Transformed GRAE and IGAE subsets, respectively. This model again generated the best f-score for both the Square Root Transformed GRAE, and IGAE subsets. After analyzing the outcomes of different perspectives, it is also shown that different classifiers provided their highest accuracy and f-score for the Z-Score Transformation and highest AUROC for Square Root Transformation methods among all outcomes. It is also observed that these classifiers gave better outcomes for IGAE subsets. Therefore, RF is considered as the most stable classifier that provides the most accurate customer purchase intention as well.

The average results of different classifiers are shown in Fig. 2. In Fig. 2a, the average values of precision, and f-measure, are slightly decreased, and the mean AUROC score is reduced compared to other metrics. After data balancing and feature engineering process, the average AUROC scores of classifiers are greatly increased (see Fig. 2b–d). In baseline, the average performances of classifiers are found better except AUROC score for CFS subsets. However, these classifiers showed better outcomes for different transformed feature subsets (i.e., especially for IGAE subsets).

6 Discussion and conclusion

Numerous e-commerce sites are competing with one another to get consumers and enhance their quality to recognize customer demands. Thus, they gather different information such as customer contacts, behaviors, responses, satisfaction, experience, etc. Many modern techniques are used to interact with customers to gather a large amount of customer records and investigate their activities. In this situation, unreliable and low-functional methods may fail because different e-commerce systems are constantly generating records that require keeping for further analysis. In this work, we proposed a machine learning model that investigates customer behavior and explores the best pipeline to detect their purchase intention more precisely. Various data balancing, feature transformation, selection, and classification algorithms were employed to conduct this work. Comparing the results of the different models, RF obtained the highest accuracy of 92.39%, f-score of 0.924 and 0.975, respectively. Many state-of-the-arts also found that RF is the best to identify customer purchasing rates as well [19, 31]. Therefore, RF is a more stable model to determine customer purchase intention more precisely. We also found Square-Root Transformation and the IGAE method as more feasible models where different classifiers are performed well than other feature engineering methods. Many works were conducted where the customer records of different e-commerce sectors were analyzed to detect their shopping intentions using various data analytics and machine learning methods [20, 22, 26, 35]. However, some of them did not consider required data cleaning, balancing, feature engineering, and machine learning methods to scrutinize online purchase intention

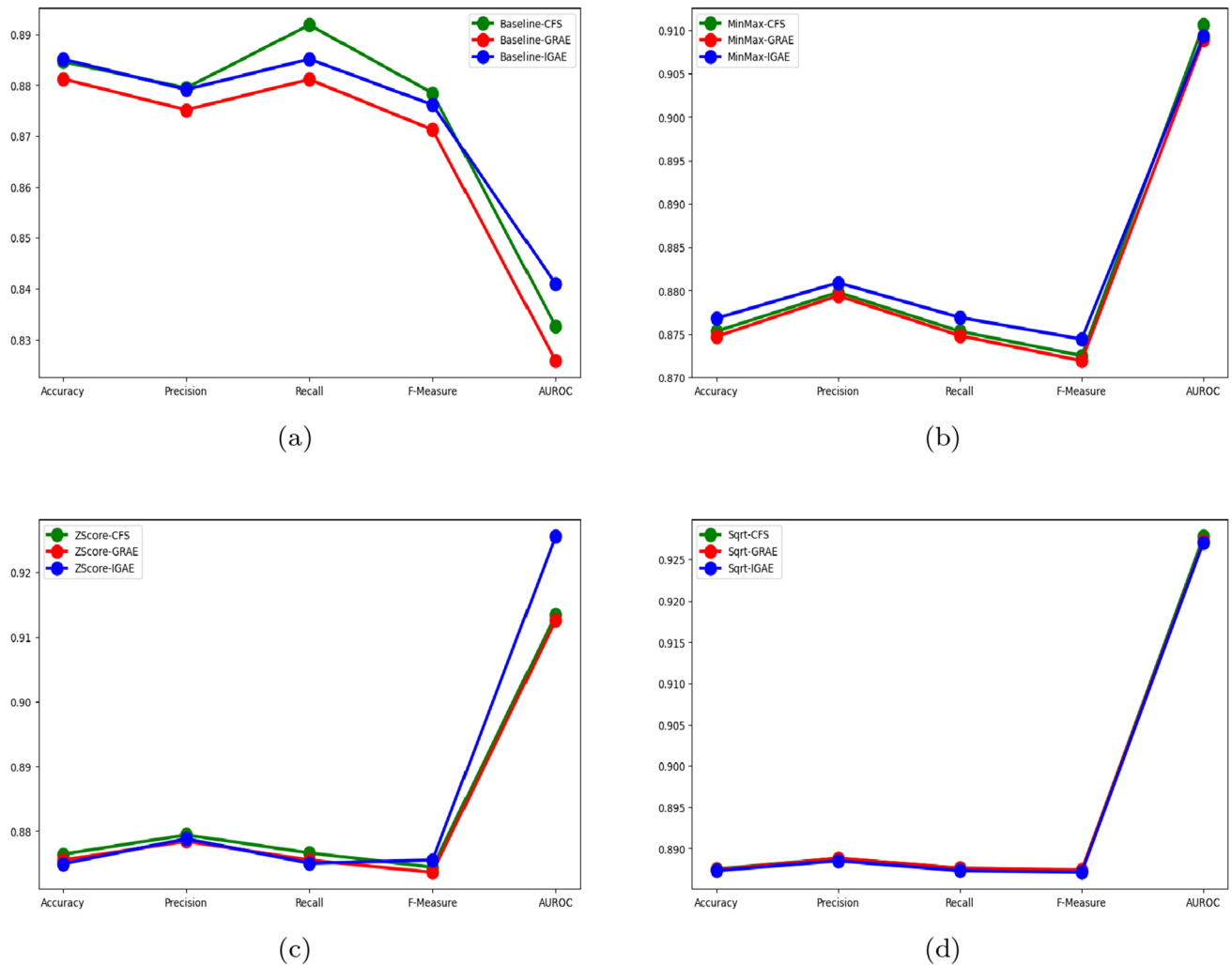


Fig. 2 Average classification results for **a** Baseline feature subsets **b** MinMax transformed feature subsets **c** Z-scored transformed feature subsets **d** Square root transformed feature subsets

[27, 34]. Several researchers also proposed a single pipeline to manipulate given customer purchasing records and did not properly compare their works with existing works [18, 20].

The implications of this model are considered from both the technical and managerial perspectives. From the technical perspective, different e-commerce systems can use this proposed methodology to get more accurate outcomes and recommend suitable and high-quality products based on customer requirements. Several significant features like product discounts, festival offers, and special/exclusive things grow customers' intention to buy these products as well. This model can extract these features to reinforce the services of e-commerce systems. Besides, this model can provide demanding products at low cost and time as well as earn huge profits from their customers. This model can also integrate with the running system to deliver extra services to the customers. From the managerial perspective, this automatic model can be taken any significant decisions about e-commerce and customer recommendation more precisely. Therefore, any type of managerial decision can be realized more quickly than traditional approaches. In the future, we will gather heterogeneous types of data from different sources and implement more advanced techniques like deep learning, transfer learning, etc. to get more suitable results for predicting customer purchase intentions in the e-commerce sector.

Author contributions This work was carried out with contributions from all authors. MSS wrote this manuscript and SFI implemented this work.

Data availability The data and materials which are generated in this work are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate This article does not contain any studies with humans and animals performed by any of the authors.

Competing interests The authors have no conflict of interest

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abedin MZ, Chi G, Uddin MM, Satu MS, Khan MI, Hajek P. Tax default prediction using feature transformation-based machine learning. *IEEE Access*. 2020;9:19864–81.
2. Abedin MZ, Hajek P, Sharif T, Satu MS, Khan MI. Modelling bank customer behaviour using feature engineering and classification techniques. *Res Int Bus Financ*. 2023;65:101913.
3. Aghdaie MH, Zolfani SH, Zavadskas EK. Synergies of data mining and multiple attribute decision making. *Procedia Soc Behav Sci*. 2014;110:767–76.
4. Allouche D, DeGivry S, Katsirelos G, Schiex T, Zytnicki M. Anytime hybrid best-first search with tree decomposition for weighted csp. In: *Principles and practice of constraint programming: 21st International Conference, CP 2015, Cork, Ireland, August 31–September 4, 2015, Proceedings 21*. Springer; 2015, pp. 12–29.
5. Apăvăloaie EI. The impact of the internet on the business environment. *Procedia Econ Financ*. 2014;15:951–8.
6. Bala M, Ali MH, Satu MS, Hasan KF, Moni MA. Efficient machine learning models for early stage detection of autism spectrum disorder. *Algorithms*. 2022;15(5):166.
7. Berk RA. Classification and regression trees (CART). In: *Statistical learning from a regression perspective*. Springer series in statistics. New York: Springer; 2008. p. 1–65. <https://doi.org/10.1007/978-0-387-77501-2>.
8. Bhagat R, Chauhan V, Bhagat P. Investigating the impact of artificial intelligence on consumer's purchase intention in e-retailing. *Foresight*. 2022;25(2):249–63. <https://doi.org/10.1108/FS-10-2021-0218>.
9. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
10. Christian TM, Ayub M. Exploration of classification using nbtrees for predicting students' performance. In: *2014 international conference on data and software engineering (ICODSE)*. IEEE; 2014, pp. 1–6.
11. Corrigan JR, Alhabash S, Rousu M, Cash SB. How much is social media worth? estimating the value of facebook by paying users to stop using it. *PLoS ONE*. 2018;13(12):e0207101.
12. Eshak MI, Ahmad R, Sarlan AB. A preliminary study on hybrid sentiment model for customer purchase intention analysis in social commerce. *2017 IEEE Conference on Big Data and Analytics (ICBDA)*. 2017, pp. 61–66.
13. Esmeli R, Bader-El-Den MB, Abdullahi H. Towards early purchase intention prediction in online session based retailing systems. *Electron Mark*. 2020;31:697–715.
14. Gomes MA, Meyes R, Meisen P, Meisen T. Will this online shopping session succeed? predicting customer's purchase intention using embeddings. *Proceedings of the 31st ACM international conference on information & knowledge management*. 2022.
15. Gupta R, Pathak C. A machine learning framework for predicting purchase by online customers based on dynamic pricing. *Procedia Comput Sci*. 2014;36:599–605.
16. Howlader KC, Satu MS, Awal MA, Islam MR, Islam SMS, Quinn JM, Moni MA. Machine learning models for classification and identification of significant attributes to detect type 2 diabetes. *Health Inf Sci Syst*. 2022;10(1):2.
17. Hussain MA, Gogoi L. Performance analyses of five neural network classifiers on nodule classification in lung ct images using weka: a comparative study. *Phys Eng Sci Med*. 2022;45(4):1193–204.
18. Islam MS, Naeem J, Emon AS, Baten A, AlMamun MA, Waliullah G, Rahman MS, Mridha M. Prediction of buying intention: factors affecting online shopping. In: *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*. IEEE; 2023, pp. 1–6.
19. Kabir MR, Ashraf FB, Ajwad R. Analysis of different predicting model for online shoppers' purchase intention from empirical data. *2019 22nd International Conference on Computer and Information Technology (ICCIIT)*. 2019, pp. 1–6.
20. Kumar A, Kabra G, Mussada EK, Dash MK, Rana PS. Combined artificial bee colony algorithm and machine learning techniques for prediction of online consumer repurchase intention. *Neural Comput Appl*. 2017;31:877–90.
21. Landwehr N, Hall M, Frank E. Logistic model trees. *Mach Learn*. 2005;59:161–205.
22. Liu C, Wang L, Lang B, Zhou Y. Finding effective classifier for malicious url detection. In: *Proceedings of the 2018 2nd international conference on management engineering, software engineering and service sciences*. 2018, pp. 240–244.

23. Liu Y, Tian Y, Xu Y, Feng Zhao S, Huang Y, Fan Y, Duan F, Guo P. Tpgn: a time-preference gate network for e-commerce purchase intention recognition. *Knowl Based Syst.* 2021;220.
24. Mu W, Lennon SJ, Liu W. Top online luxury apparel and accessories retailers: what are they doing right? *Fash Text.* 2020;7(1):1–17.
25. Noviantoro T, Huang JP. Applying data mining techniques to investigate online shopper purchase intention based on clickstream data. *Rev Bus Account Financ.* 2021;1(2):130–59.
26. Sakar CO, Polat S, Katircioglu M, Kastro Y. Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and lstm recurrent neural networks. *Neural Comput Appl.* 2019;31:6893–908.
27. Sang G, Wu S. Predicting the intention of online shoppers' purchasing. 2022 5th International conference on advanced electronic materials, computers and software engineering (AEMCSE). 2022, pp. 333–337.
28. Satu MS, Howlader KC, Barua A, Moni MA. Mining significant pre-diabetes features of diabetes mellitus: a case study of Noakhali, Bangladesh. In: *Applied informatics for industry 4.0*. Chapman and Hall/CRC;2023, pp. 280–292.
29. Satu MS, ZoynulAbedin M, Khanom S, Ouenniche J, ShamimKaiser M. Application of feature engineering with classification techniques to enhance corporate tax default detection performance. In: *Proceedings of international conference on trends in computational and cognitive engineering: Proceedings of TCCE 2020*. Springer; 2021, pp. 53–63.
30. ShahriareSatu M, Atik ST, Moni MA. A novel hybrid machine learning model to predict diabetes mellitus. In: *Proceedings of international joint conference on computational intelligence: IJCCI 2019*. Springer; 2020, pp. 453–465.
31. Shi X. The application of machine learning in online purchasing intention prediction. *Proceedings of the 6th international conference on big data and computing.* 2021.
32. Siers MJ, Islam MZ. Cost sensitive decision forest and voting for software defect prediction. In: *PRICAI 2014: trends in artificial intelligence: 13th Pacific Rim international conference on artificial intelligence, Gold Coast, QLD, Australia, December 1–5, 2014*. Proceedings 13. Springer; 2014, pp. 929–936.
33. Sunny FA, Khan MI, Satu MS, Abedin MZ. Investigating external audit records to detect fraudulent firms employing various machine learning methods. In: *Proceedings of the Seventh International Conference on Mathematics and Computing: ICMC 2021*. Springer; 2022, pp. 511–523.
34. Trivedi SK, Patra P, Srivastava PR, Zhang JZ, Zheng LJ. What prompts consumers to purchase online? A machine learning approach. *Electronic Commerce Research.* 2022;pp. 1–37.
35. Zheng B, Liu B. A scalable purchase intention prediction system using extreme gradient boosting machines with browsing content entropy. 2018 IEEE International Conference on Consumer Electronics (ICCE). 2018, pp. 1–4.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.