Customer Behavior Analysis using Wild Horse Optimization Algorithm

Raheleh Sharifi[1], Mohammadreza Ramezanpour[2]
1- Department of Computer Engineering, Majlesi Branch, Islamic Azad University, Isfahan, Iran.
Email: r.sharifi1365@yahoo.com
2- Department of Computer Engineering, Mobarakeh Branch, Islamic Azad University, Isfahan, Iran.
Email: ramezanpour@mau.ac.ir (Corresponding author)

Received: 7 March 2023 Revised: 20 April 2023 Accepted: 15 May 2023

ABSTRACT:
One of the areas in which businesses use artificial intelligence techniques is the analysis and prediction of customer behavior. It is important for a business to predict the future behavior of its customers. In this paper, a customer behavior model using wild horse optimization algorithm is proposed. In the first step, K-Means algorithm is used to classify based on the features extracted from the time series, and then in the second step, wild horse optimization algorithm is used to estimate customer behavior. Three datasets including, the grocery store dataset, the household appliances dataset, and the supermarket dataset are used in the simulation. The best clusters count for the grocery store dataset, the household appliances dataset, and the supermarket dataset are obtained 5, 4, and 4, respectively. The simulation results indicate that this proposed method is obtained the lowest prediction error in three simulated datasets and is superior to other counterparts.

KEYWORDS: Customers' Behavior Analysis, Clustering, Time Series Features, Wild Horse Optimization.

1. INTRODUCTION

Today, one of the biggest challenges of organizations, institutions and companies around the world is to identify customers and the profitability of their different groups [1]. The importance of this issue is much greater for stores that are in contact with different groups of customers [2]. Since the resources of these stores are limited; it is better to make it available to profitable customers and face with the increase of profits in their business [3, 4]. In this regard, researchers have identified different groups of customers by presenting several models and algorithms.

One of the most popular models has been discussed in this regard and has attracted the attention of all researchers is the customer relationship management model [5] Customer relationship management is a concept or strategy to strengthen relationships with customers while costs are reduced and productivity and profitability in business are increased [6, 7]. An ideal customer relationship management system is a collection of all data sources of an organization that provides the information needed by the customer in a short time [8]. The customer relationship management system is extensive and interesting, but it can be implemented for small businesses as well as large companies to help customers [7, 9-11].

Customer relationship management covers various activities. This model helps the managers of companies and industries in adopting important strategies by using the knowledge of data mining and extracting the hidden knowledge of raw data [10]. Data mining is defined as the process of extracting and discovering hidden patterns from big data. Data mining tools have been widely used in the field of customer relationship management [12]. Predicting and clustering are data mining techniques that are widely used to analyze and forecast customer behavior [13]. Especially, customer segmentation is a popular approach that divides customers into groups with similar needs, characteristics and behaviors [14].

The literature related to customer relationship management shows that most of the studies conducted for segmentation have used a static approach, while one of the most important challenges in customer segmentation is emphasis on the dynamics of behavior, because several indicators such as market changes, product promotion, new product introduction, and competitors' approaches change customer behavior over time [15]. Therefore, the dynamic nature of customer behavior must be considered for the effectiveness of the customer segmentation process [16]. Accordingly, the

Paper type: Research paper
DOI: 10.30486/MJTD.2023.1980070.1025
How to cite this paper: R.Sharifi, M, Ramezanpour, "**Customer Behavior Analysis using Wild Horse Optimization Algorithm**", *Majlesi Journal of Telecommunication Devices*, Vol. 12, No. 2, pp.79-93, 2023.

proposed plan will focus on the dynamic nature of customer behavior.

Feature extraction and time series analysis are two main approaches to analyze the dynamic behavior of customers [15]. It will be tried to discover the hidden knowledge in the time series data of customers using the clustering approach by considering these approached in the proposed plan. Each customer in the proposed model can be a member of one of the clusters according to his/her characteristics and behavior (characteristics extracted from the time series). In the following, customer behavior is estimated based on an efficient predictive algorithm. The proposed algorithm for estimating customer behavior is the improved decision tree algorithm with the wild horse optimization algorithm [17]. The algorithm that will be called DTWHO in this research.

The wild horse optimization algorithm was presented in 2021 and its performance in continuous and discrete problems confirmed that its performance is better in comparison with similar optimization algorithms and it is less trapped by local optimum. In the DTWHO algorithm, the predictive parameters of the decision tree are adjusted with a meta-heuristic approach so that the algorithm has its best performance so that the prediction error reaches the minimum possible value.

2. BACKGROUND

This article is a combination of several data mining approaches to analyze customer behavior. It is necessary to be familiar with some of the concepts used in the article in order to understand better the content of the text. Concepts related to the proposed topic are described in the rest of this section.

2.1. Customer behavior

Customer behavior includes actions, processes and social communication that are carried out by individuals and groups before, during, and after the exchange process. Customer behavior includes processes that are effective in choosing, consuming, and buying goods and services. These processes include emotional, mental and behavioral factors of customers. Customer behavior includes various steps that a customer goes through before buying the products or services used by him/her [18]. Customer behavior is the study of the processes which are used to choose, use (consume) and dispose products and services, including emotional, mental and behavioral reactions of consumers. In fact, customer behavior includes ideas from several sciences, including psychology, biology, chemistry and economy [19].

2.2. Regression

Regression is a well-known statistical learning technique that is used to infer the relationship between the dependent variable Y and the independent variable P of [X1...XP]. Precisely, regression analysis tries to use statistical observations to determine Y according to X as Y=fX [20]. The researcher seeks to determine the causal effect of one variable on another variable in regression analysis; for example, investigating the impact of price increases on demand, or investigating the impact of changes in money supply on inflation [21]. Regression analysis techniques can be divided into two main categories without parameters and parametric. In the first category, a specific form for f has not been determined. Actually, the form of the functional relationship between the dependent variable and the independent variable is not predetermined. While the second category includes those techniques that work by assuming a relationship with a number of Beta parameters that must be estimated using the observed dataset [20].

2.3. Decision tree

A decision tree is a decision support tool that uses trees for modeling. The decision tree is commonly used in various applications of machine learning. The application of the decision tree is to describe the calculation of conditional probabilities. The decision tree is a supervised learning algorithm and is used to classify data. A decision tree uses a tree to build a prediction (estimation) model that maps observations about a segment to conclusions about a target value for that segment. In the structure of the trees, the leaves represent the class labels and the branches represent the relationship of the features that branch to these labels. In this study, the CART decision tree is used. This method, which leads to the formation of a decision tree with binary divisions, was fully introduced by Beriman et al. in 1984 [22].

2.4. Wild horse optimization algorithm

The wild horse optimization algorithm is a crowd-based optimization algorithm that was introduced in 2021. This algorithm models the behavior of wild horses in grazing and mating. Wild horses have a hierarchical life that horses are divided into territorial and non-territorial groups in this structure. These groups have many differences in terms of grazing, mating, leadership, hierarchy and dominance. The focus of the wild horse optimization algorithm is on non-territorial horses. Non-territorial horses are herds with one stallion and one or more mares and offspring. Sometimes, single males (adult) and male colts (immature next to mares) are observed in the herd of non-territorial horses. Colts usually start the grazing process in the first week of life and graze most of the time. When colts grow they rest less. The colts in this colony leave their parents before puberty and the male colts (close to puberty) join single groups to be ready for mating. Female foals also join family groups. The departure of male colts near puberty is to prevent father-daughter mating or sibling mating. In this colony, each group has a leader. The leader of the group is usually the most dominant horse and is a stallion; the rest of the group follows the leader in the order of decreasing dominance (position in the hierarchy). In the optimization algorithm of wild horses, the effort is to provide an efficient meta-heuristic approach for optimization problems from examining group behaviors, grazing, mating, dominance and leadership of non-territorial horses. A five-step structure for wild horse optimization algorithm can be considered by studying a colony of non-territorial horses. Based on this structure, groups are formed after creating the initial population and leaders are chosen for each group. In the continuation of the process of grazing and mating of horses, the stallions lead the groups. Next, the process of exchanging and choosing the leader of the groups is done. In this algorithm, the position of the best horse in terms of fitness is stored for each generation. Examining the performance of the wild horse optimization algorithm in 13 fitness functions shows that this algorithm is less trapped in local optimum in comparison with other similar algorithms such as Whale, Gray Wolf, Genetics, etc. and produces better answers. The following assumptions are valid for the wild horse optimization algorithm.

- The search of the problem space is guaranteed by the random selection of the leader of the groups and the random movement of the colts around the leader of each group.
- Due to the departure of adult horses from the group and mating with adult horses of other groups, this probability will be less that algorithm is trapped in local optimum.
- The wild horse optimizer is a crowd-based algorithm.
- The random movements of colts are to create diversity in the population.
- The group leader is moved towards the optimum during the optimization process.
- The leader of the groups leads the horses to the desired areas in the problem space.
- The best leader in each iteration is saved and compared with the best leader obtained up to that moment (optimum).
- The wild horse optimizer algorithm has very few parameters to adjust.
- The wild horse optimizer algorithm is non-gradient.

3. RELATED WORKS

In [1], a simple and efficient method based on homogenous feature selection called bagging homogeneous feature selection (BHFS) is proposed for analyzing customer behavior. The BHFS method is used to remove irrelevant features from the dataset and select a stable feature subset and to improve the predictive power of the NB algorithm. The BHFS method requires less execution time and selects the best relevant features for evaluation. The experimental results show that the model based on BHFS makes better predictions when is compared to the standard NB. The runtime complexity is also lower with BHFS-NB.

The article [2] investigates the behavior of customers (Vietnamese and non-Vietnamese hotels) with focused on five features (service, value, room quality, sleep quality and cleanliness). The dataset includes the feedbacks of 1357 customers from 70 countries and five continents that have been collected in 467 hotels and six tourist cities in Vietnam. The proposed approach for analysis of customer behavior is descriptive statistics, t-test and one-way analysis of variance. Examining the results shows that complaints about service and value are more evident among Vietnamese customers. While non-Vietnamese customers complained about the condition of the rooms. Also, being dissatisfied with the service, cleanliness, room and sleep quality is often more common among guests of economy class hotels.

The article [3] has investigated whether customers behave differently in the transactions of a supermarket in Taiwan before and during the Chinese New Year holiday. The results show that customers visit the supermarket the most during the Chinese New Year holiday. And customers who shop before the Chinese New Year holiday; On average, they make more purchases. The proposed approach of this article for analyzing customers is based on statistical methods and variance analysis. Although these approaches provide more tangible results; but they do not discover the hidden knowledge of raw data.

In [4], researchers have presented a model for analyzing the behavior of customers of an electronic product store and a food company by relying on data mining models. In this article, cumulative clustering is a type of hierarchical clustering; It is suggested to group customers.

The article [5] has proposed a two-dimensional approach to analyze the behavior of electronic and traditional banking customers in Iran. For this purpose, one-year transactions of customers of one of the largest private banks in Iran (by consisting one million records) were collected and the data were collected using RFM model (recentness, frequency and financial ability) were clustered to extract the hidden knowledge of the data using the CRISP-DM method and the K-Mean algorithm was used for clustering. Also, the customer lifetime value was calculated using Lawshe's criteria. Based on the results, customers who use electronic and traditional banking services have more value.

In [6], researchers have proposed an intuitive fuzzy clustering approach to analyze customer behavior. In order to model their proposed method, they have collected the transactions of two major cities in Turkey, which are more than 10,000 records, and implemented their plan. Examining the results shows that the intuitive fuzzy clustering approach performed better than similar algorithms in terms of clustering efficiency. Although the proposed approach provides better results, it should be kept in mind that methods based on fuzzy inference are dependent on the primary center of clusters and defined fuzzy rules.

In [7], a new method for analyzing customer behavior is presented by relying on the generation of multivariate time series from user data and determining the similarity between them. In this article, the time series of each user is modeled based on a set of features and over time. Researchers have used the score of the similarity matrix to measure the effectiveness of the proposed model. This index creates a matrix that includes the similarity of the time series of all users. Results show that the model proposed in this article provides better results compared to similar works; but it should be considered that it also has a high computational complexity.

The main purpose of [8] is to analyze the behavior of customers in front of different products. The analysis process of the article includes three main stages of calculating the quality, the customer, calculating the quality of the customer's behavior and comparing the results. In this study, Chi-Square is used to analyze customer behavior. The findings proved that the quality of the product is different depending on the customer's profile. In addition to this, the quality of the customer has a direct relationship with the customer's behavior of the products. Another analysis of this article to measure the willingness to buy the product among customers showed that the brand name and price have an effect on the buying behavior of customers. However, the effect of factors such as product quality, social networks and customer profiles on the purchasing behavior of customers should not be ignored.

In [9], researchers have used binary logistic regression to investigate the buying behavior of consumers of agricultural products. The results show that gender, age, consumer perception of the product and monthly income have a positive effect on the stability of behavior and willingness to buy of consumers; while the price of agricultural products and online shopping have a negative effect on the willingness of consumers.

The article [10] has presented a new approach by relying on deep learning and customer behavior analysis to detect fraud in financial transactions. Because the dataset is the imbalance in this article, a solution to overcome this challenge is presented, first; then, a new model is used to analyze customer behavior by combining convolutional network and long-term short-term memory network. The results show that the efficiency of the proposed method is better in comparison with other deep learning models, and fraudulent behaviors can be recognized better.

In [25], researchers have presented a model for analyzing the behavior of customers of an electronic product store and a food company by relying on data mining models. In this article, cumulative clustering, which is a type of hierarchical clustering, is proposed for grouping customers.

The main purpose of [29] is to analyze the behavior of customers in front of different products. The analysis process of the article includes three main stages of calculating the quality, the customer, calculating the quality of the customer's behavior and comparing the results. In this study, Chi-Square is used to analyze customer behavior. The findings proved that the quality of the product is different depending on the characteristics of the customer. In addition, the situation or the quality of the customer has a direct relationship with the customer's behavior of the products. Another analysis of this article to measure the willingness to buy the product among the customers showed that the brand name and price affect the buying behavior of customers; however, the effect of factors such as product quality, social networks and customer profiles on the buying behavior of customers should not be ignored.

In the article [33], instead of clustering on the time series itself, first a number of features were extracted from the time series based on the features of [34]; then, clustering was done based on the extracted features. K-means, K-medoids and DBSCAN methods were used for clustering. And they have used LSTM method in the prediction part.

4. THE PROPOSED METHOD

Various approaches have been proposed to analyze the behavior of customers, among which one of the approaches that has received more attention is the designs based on data mining algorithms and machine learning. It will be tried in the proposed plan to provide an efficient algorithm for analyzing customer behavior by combining clustering and forecasting methods. Customer behavior is modeled in the form of time series. The proposed plan includes two general steps. The customers are separated into clusters based on the features extracted from the time series in the first step, and the customers' behavior is estimated based on an efficient predictive algorithm in the second step. The datasets are first converted into a standard format by the pre-processing process to achieve the outlined goals; then, the first step which is clustering is implemented and the customers are separated into different clusters based on the features extracted from the time series. A decision tree is a decision support tool that uses trees to model. A decision tree is typically used in various machine learning applications. A decision tree is a decision support tool that uses trees to model. A decision tree is typically used in various machine learning applications. A decision tree can be used to describe conditional probability calculations. A decision tree is a supervised learning algorithm and it is used for classification of data and also for prediction. Decision tree uses a tree to build a prediction model (estimation) that maps observations about a part to conclusions about the target value of that part. The leaves in the tree structure represent class tags and branches indicate the relationship of features that are branched to these tags. The decision tree like other machine learning algorithms has parameters whose values are set by default. More promising results can be obtained by fine adjustment of the parameters of this algorithm with respect to the dataset on which the machine learning process is performed, but adjustment of the parameters of the decision tree algorithm is an NP-problem. It is recommended to use optimization algorithms to solve such problems. Accordingly, decision tree parameters are used for predicting customer behavior in the proposed plan by use of a new optimization algorithm that is less likely to be trapped in the local optimum and has favorable convergence. It is found based on the research that the wild horse optimization algorithm [17] can be a suitable tool for this purpose. The wild horse optimization algorithm was presented in 2021 and its performance analysis in continuous and discrete problems confirms that its performance is better compared to similar optimization algorithms and it is less trapped in local optimums. Customer behavior prediction is performed on the representative vectors of the clusters and customer behavior is predicted, and the prediction error is calculated for each cluster with each of the methods by using a known error evaluation criterion. Algorithm (1) shows the general structure of the proposed plan approach for analyzing customer behavior.

4.1. First step: clustering

Customer will be clustered based on time series features. For this purpose, pre-processing operations are performed on the datasets, first. Then, the extracted features of time series (the customers' time series) and the ranking of the features are done. The data are also clustered at the end of this step. The feature extraction process in the proposed plan is performed based on the approach of article [33]. Table (1) introduces the list of features that are extracted in this step.

Table 1. Introducing features extracted from time series in the proposed method

Feature name	Explanation
mean	Mean of all time series points
Var	Variance of all time series points
ACF1	First order correlation
Trend	Trend strength in time series
Linearity	The degree of linearity of the time series
Curvature	The degree of curvature of the time series
Entropy	Spectral entropy - Shannon entropy
Lumpiness	Variance change in residuals
Spikiness	The degree of stickiness of the time series
Lshift	Change of the level using the sliding window
Vchange	Variance change
Fspots	Number of planar points using discretization
Cpoints	The number of times that a time series crosses the median

Algorithm 1. Pseudo-code of wild horse optimization algorithm

- Dataset Preprocessing and informing time series for each sample
- Features extraction from time series of each sample (15 features)
- Ranking and features selection (reducing computational complexity)
- Features Normalization
- Dataset clustering (based on features selection)
- Calculating the mean time series for each sample in the clusters
- Introducing the representative of each cluster
- Customer behavior analysing using improved decision tree algorithm
  - Initialize population of horses randomly (for each cluster)
  - % Solutions for decision tree parameters setting %
  - Input WHO parameters, PC=0.13, PS=0.2
  - Calculate the fitness of Horses (Customer behavior prediction error)
  - Create Colt groups and select Stallions
  - Find the best Horse as the optimum (horse or solution with least amount of error)
  - While the end criterion is not satisfied
  - Calculate TDR
    - For number of Stallions
      - Calculate Z (An adaptive mechanism)
        - For number of Colts of any group
          - If rand>PC
            - Update the position of the Colt (Using Grazing behaviour)
          - Else
            - Update the position of the Colt (Using Horse mating behavior)
          - End
        - End
      - If rand>0.5

- Update the position of the $\overline{\text{Stallıon}_{G_\iota}}$(Using Group leadership - F1)
- Else
  - Update the position of the $\overline{\text{Stallıon}_{G_\iota}}$ (Using Group leadership - F2)
- End
- If cost ($\overline{\text{Stallıon}_{G_\iota}}$) < cost(Stallion)
  - Stallion = $\overline{\text{Stallıon}_{G_\iota}}$
- End
- sort Colts of group by cost
- select Colt with Minimum cost
- If cost(Colt) < cost (Stallion)
  - Exchange Colt and Stallion Position (Using Group leadership)
- End
- End
- Update optimum (horse or solution with least amount of error)
- % A solution for setting decision tree parameters %
- End
- Using optimum to decision tree parameters setting
- % Improved decision tree %
- Using an improved decision tree in customer behavior analysis

The extracted features are ranked and n valuable features are selected for the feature selection process in the following. Feature selection is done with the aim of reducing the time complexity of the proposed model. The proposed approach of this article for this step is the Laplacian scores method [35]. The extracted features are ranked based on the score they receive by the Laplacian scores method. The samples in the dataset are clustered based on the selected features. Clustering is one of the branches of unsupervised learning and is an automatic process during which samples are divided into categories whose members are similar to each other. These categories are called clusters. Therefore, a cluster is a collection of objects in which the objects are similar to each other and are not similar to the objects in other clusters [36]. After clustering the samples (based on the valuable features) in the proposed plan, the time series of the samples is called and their means are calculated for each cluster. The resulting mean is considered as the representative of the cluster.

4.2. Second step: Predicting customer behavior

In the second step of proposed plan, the representatives of each cluster are received as input using the output of the first step which is based on clustering, and the dataset of the second phase is constructed using it. The sliding window approach is used to construct the second step dataset. Accordingly, the data related to each cluster is generated after determining the length of the window and the length of the step. The most important goal of the second step is to predict the representative for each cluster with the lowest amount of error. Various algorithms have been introduced for the prediction process, but various studies show that machine learning algorithms have better results than statistical methods [37]. There are different types of machine learning algorithms; one of the most efficient algorithms for that is the decision tree. Although this algorithm provides effective results in machine learning, it should be considered that its high efficiency depends on the optimal setting of parameters. The machine learning process with the decision tree in the default mode starts by the initial values of the parameters and the same settings are maintained until the end. Therefore, it cannot be claimed that better accuracy can be achieved by the default values of the parameters which are random and based on little experiences [38]. Thus, it is necessary to adjust the parameters of the decision tree according to the conditions of the problem and the dataset in order to achieve better results. But it is an NP-hard problem: what are the values of the parameters to reduce detection error [39]. Since the use of meta-heuristic optimization algorithms is considered a suitable approach to solve NP-Hard problems [40]; it can be said that updating and adjusting the values of the decision tree parameters can be presented in the form of an optimization problem. Table (2) shows the variables considered in the decision tree.

Table 2. Decision tree parameters to adjust in the proposed model

Row	Parameter	Definition	Value
1	PredictorSelection	predictor selection	'allsplits','curvature', 'interaction-curvature'
2	Prior	Priority	'empirical','uniform'
3	MaxNumSplits	Maximum number of splits	Integer
4	MinLeafSize	Minimum leaf size	Integer
5	NumVariablesToSample	Number of selected variables	Integer
6	SplitCriterion	Criteria for pruning branches	‘gdi’,'twoing', 'deviance'

The initial population of the wild horse optimization algorithm in the proposed plan is equal to the generation of random solutions to adjust the parameters of the decision tree. These solutions are scattered in the problem space (search space) and each one will represent a solution. Therefore, it can be said that the position of wild horses in the proposed method is a potential solution for adjusting the parameters of the decision tree, but if the mathematical modeling of the vectors of wild horses is used in the proposed plan, the solution of the i-th wild horse can be shown as a vector $\mathrm{x_i}$, which is modeled according to Eq. (1).

$$\mathrm{x_i} = [x_i^1, \mathrm{x_i^2}, \dots, \mathrm{x_i^{nv}}] \tag{1}$$

In this modeling, $\mathrm{X_i^j}$ represents the j-th parameter in the decision tree from the i-th wild horse and nv represents that the length of the wild horse vector is equal to the result of the parameters of the decision tree. It is intended to adjust 6 parameters of the decision tree in the proposed plan; therefore, the vector length of each wild horse will be equal to 6. By modeling wild horses, the initial population of wild horse optimization algorithm can also be mathematically modeled.

$$\mathrm{P}_{WHO} = \begin{bmatrix} \mathrm{W_{1,1}} & \dots & \mathrm{W_{1,d}} \\ \dots & \dots & \dots \\ \mathrm{W_{n,1}} & & \mathrm{W_{n,m}} \end{bmatrix} \tag{2}$$

The initial population will be randomly divided into several groups after the generation of the initial population according to the colony of non-territorial (wild) horses. If the parameter N is the number of members of the population (herd); the number of groups will be $G = \lceil N \times PS \rceil$. In this regard, the PS parameter is the percentage of stallions in the total population of horses. This parameter is one of the inputs of the wild horse optimization algorithm. Therefore, there are G leaders (male horse) in the wild horse optimization algorithm, according to the number of groups and the remaining members of the initial population $(N - G)$ are equally divided between the groups.

Then, the fitness of each member of the initial population is calculated and leaders (the best fitness value) are selected among the group members based on the obtained fitness. The fitness of wild horses in the proposed method can be modeled as a matrix and stored in each generation of the algorithm. The relation (3-4) introduces this modeling. The created matrix has one row and as many columns as wild horses.

$$M_f = \begin{bmatrix} f(w_{1,1}, w_{1,2}, \dots w_{1.m}) \\ \dots \\ f(A_{n,1}, A_{n,2}, \dots, A_{n,m}) \end{bmatrix} \tag{3}$$

Each line in relation (3) will represent the fitness of each of the wild horses, each of which is a solution for setting the parameters of the decision tree. The purpose of the proposed method in calculating the fitness of horses is to reduce the prediction error rate by the decision tree, which is modeled according to Eq. (3). This criterion calculates the prediction error rate of a predictive algorithm. For this purpose, the data are divided into two categories: training and testing (80% of the samples for training and 20% of the samples for test) for each of the solutions proposed by wild horses. The decision tree is adjusted according to the proposed model of wild horse and passes the training step by use of the training data. Then, its efficiency is tested by using the test set. The prediction error obtained at this step is considered as fitness for each wild horse. Eq. (4) shows the fitness function of wild horses.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(t_i - y_i)^2}{N}} \tag{4}$$

where the N represents the number of samples, the t_i represents the actual value for the i-th sample, and the y_i represents the predicted value for the i-th sample.

The next step in the second step of the proposed plan is the modeling of grazing behavior based on the colony structure of non-territorial horses. Colts usually spend most of their time for grazing around their group. The stallion is considered in the center of the grazing area in order to model this behavior, and the rest of the group members perform the search operation (grazing) around the stallion. Eq. (5) is proposed to simulate the grazing behavior. This equation expresses the movement of group members with different radii around the group leader (stallion) to find better forage (search).

$$\bar{X}_{i,G}^{j} = 2Z\cos(2\pi RZ) \times \left(\mathrm{Stallion}^j - X_{i,G}^j\right) + \mathrm{Stallion}^j \tag{5}$$

where $x_{i,G}^i$ are the current position of the group member (stallion or mare), $Stallion^i$ is the position of the stallion (group leader) and Z is an adaptive mechanism (Z is calculated by Eq. (6), R is a uniform random number in the range of [-2, 2] (this parameter causes the horses to graze at different angles relative to the group leader), the π is the pi number equal to 3.14. Also, the COS function in this equation, with the combination of π and R causes the movement of horses in a radius becomes different . And finally, $x_{i,G}^i i$ is the new position of the group member, when it is grazing.

$$P = \vec{R}_1 < \mathrm{TDR}; \quad IDX = (P == 0); \quad Z = R_2 \Theta \mathrm{IDX} + \vec{R}_3 \Theta (\sim \mathrm{IDX}) \tag{6}$$

where P is a vector consists of 0 and 1 equal to the dimensions of the problem. $\overrightarrow{R_3}$ and $\overrightarrow{R_1}$ vectors are random vectors with a uniform distribution in the range of [0, 1]. Also, R_2 is a random number with a uniform distribution in the range of [0,1]. IDX indicators generate random vector $\overrightarrow{R_1}$ that satisfy the condition of (0==P). The TDR is also an adaptive parameter that starts with a value of 1 and decreases during the run of the algorithm according to Eq. (7); this parameter reaches 0 at the end of the algorithm run.

$$\mathrm{TDR} = 1 - \mathrm{iter} \times \left(\frac{1}{\mathrm{maxiter}}\right) \quad (7)$$

where *iter* is the current iteration and *maxiter* is the maximum number of iterations of the algorithm.

Another wild horse operator which is also relied on in the proposed plan is the mating operator. One of the unique behaviors of horses compared to other animals is the separation of colts from the group. In this process, the colts leave the group before reaching maturity. Based on this process, male colts join single groups and female foals join other groups to find their mate. This departure is to prevent mating father with his daughter or mating siblings with each other. This behavior is simulated in this way that one colt leaves the group *j* and joins to a temporary group. It is assumed that the temporary group includes one colt and one foal. Since these two colt and foal have no family relationship; they can mate and have children after puberty, but their child must leave the temporary group and go to another group like *k*. This cycle of exit, mating and reproduction is repeated for all different groups. Eq. (8) is proposed to simulate the behavior of exit and the mating of horses. This equation is the mean-type crossover operator.

$$X^{p}_{G,K} = \mathrm{Crossover}(X^{q}_{G,i}, X^{z}_{G,j}) \quad i \neq j \neq k, p = q = \mathrm{end}. \quad (8)$$
$$\mathrm{Crossover} = \mathrm{Mean}$$

where $X^{P}_{G,K}$ represents the position of horse p (a horse that has reached the age of puberty and must leave the group) from group k, which leaves the group and gives its place to a horse whose parents are in groups i and j. Also, $X^{q}_{G,i}$ is the position of colt q from group i, which exits from group and it is mated after puberty by horse z with the position of $X^{Z}_{G,j}$ that exits from group z.

The group leader in a colony of non-territorial horses must lead the group to an appropriate area. The appropriate area is considered in the optimization algorithm of the wild horses as a water hole. The group should move towards the water hole; while another group is moving towards the same water hole. The leaders of the groups compete to take over the water hole. A group that has more dominance (dominant group) than other groups will use the water hole. In this modeling, other groups are allowed to use the water hole when the dominant group moves away from the water hole. Eq. (9) is proposed to perform this operator in the wild horse optimization algorithm.

$$\overline{\mathrm{Stallion}_{G_i}} = \begin{cases} 2Z\cos(2\pi RZ) \times (\mathrm{WH} - \mathrm{Stallion}_{G_i}) + \mathrm{WH} & \text{if } R_3 > 0.5 \\ 2Z\cos(2\pi RZ) \times (\mathrm{WH} - \mathrm{Stallion}_{G_i}) - \mathrm{WH} & \text{if } R_3 \leq 0.5 \end{cases} \quad (9)$$

where $\overline{Stallion_{G_i}}$ is the next position of the leader of group *i*, the *WH* is the position of the water hole, $Stallion_{G_i}$ is the current position of the leader of group *i*, *Z* is an adaptive mechanism calculated by Eq. (9), R is a uniform random number in the range of [-2, 2], the π is Pi number and equal to 3.14.

The group leaders are randomly selected at the beginning of the wild horse optimization algorithm to maintain the random nature of the algorithm. But, the leaders are selected based on the fitness value in the next steps of algorithm. If the fitness of a member of the group is better than the leader of that group, the position of the leader and the corresponding member will change according to Eq. (10).

$$\mathrm{Stallion}_{G_i} = \begin{cases} X_{G,i} & \text{if } \cos t(X_{G,i}) < \cos t(\mathrm{Stallion}_{G_i}) \\ \mathrm{Stallion}_{G_i} & \text{if } \cos t(X_{G,i}) > \cos t(\mathrm{Stallion}_{G_i}) \end{cases} \quad (10)$$

5. ANALYSIS

The proposed method is simulated in the MATLAB 2016 software environment and the Symmetric Mean Absolute Percentage Error (SMAPE) index is used to evaluate and compare the results with similar works. Eq. (11) formulates this index [41].

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|\hat{y_t} - y_t|}{\frac{|\hat{y_t}| + |y_t|}{2}} \quad (11)$$

In (11), the y_t represents the real value and $\hat{y_t}$ represents the predicted value at the time point t.

The simulation process of the proposed model is implemented on three datasets of grocery and supermarket and household appliances. The research dataset consists of 11-month POS device data collected from different stores. The total number of terminals is 12,3684, among them, 60,159 terminals which have regular transactions during 44 weeks. Among these terminals, we selected 416 terminals related to household appliance stores, 257 terminals related to grocery stores, and 10,578 terminals related to supermarket stores, that the third quarter is selected from supermarket data which includes 2644 terminals. Table (3) introduces the generalities related to these datasets.

Table 3. Introducing the datasets used in the research about the number of terminals

Dataset title	Number of terminals	Number of transactions	Number of clusters
Household appliances	416	44 series per terminal	4
Grocery	257	44 series per terminal	5
Super Market	2644	44 series per terminal	4

5.1. Selection of effective features

The clustering process is done in the proposed plan only by emphasizing on the effective features in order to reduce the time complexity. The criteria for ranking features in the proposed article is the Laplacian scores. Table(4) shows the results of this survey.

Table 4. Feature score based on Laplacian approach

Feature	Laplace score
mean	0.9605
Var	0.9942
ACF1	0.9684
Trend	0.9755
Linearity	0.9611
Curvature	0.9547
Entropy	0.9795
Lumpiness	0.9797
Spikiness	0.9851
Lshift	0.9625
Vchange	0.9681
Fspots	0.9794
Cpoints	0.9510

Focus on 4 features is effective in the proposed plan, like basic plan; and the clustering process is done by focus on four effective features. According to the results of the Laplacian test, 4 features of Var, Spikiness, Lumpiness and Entropy are the most effective features of the dataset and the clustering process will be based on them.

5.2. Clustering

The process of clustering the samples begins after pre-process of the dataset and selection of the effective features. For this purpose, it should be checked each terminal with how many clusters is suitable for analyzing customer behavior. In [23], an overview is done for this purpose. The researchers are performed the clustering process for each terminal with 4, 5, 6, 7 and 8 clusters. The results of this study shown in Table (5) indicates that the best case for the household appliance dataset, the best case is with 4 clusters, and for the supermarket dataset, the best case is with 4 clusters. The results of [23] are documented in this article.

Table 5. Results of customer clustering in different datasets

	method/ number of clusters	4	5	6	7	8
Grocery	K-means	0.5768	0.6213	0.5745	0.5675	0.5640
Household appliances	SOM	0.7622	0.6155	0.5411	0.5600	0.4940
Super Market	K-medoids	0.6378	0.5616	0.5439	0.5011	0.5269

5.3. Predicting customer behavior

In this section, the performance of the proposed method has been evaluated on three datasets of household appliances and grocery stores by the SMAPE index, and the results are compared with similar works under similar conditions. The results of the first step in the proposed clustering plan show that the number of representative vectors of the clusters (in the best possible case) is 4 for the supermarket set, 4 for the household appliances dataset, and 5 for the grocery store dataset. In this step, each representative vector is converted into a dataset by using the sliding window and it is used for prediction. In this experiment, the length of the time window is considered as a control parameter by taking into account that the length of the sliding window is effective on the samples and also the number of features and different interpretations reach by changing it. For this purpose, tests will be performed for different lengths of the time window (1 to 6).

a) Predicting consumers' behavior of household appliances

In this section, the effectiveness of the proposed model (DTWHO) in predicting customers' behavior is investigated. This experiment is performed for different lengths of the time window (1 to 6) and the results are analyzed. Table (6) shows the results of this investigation, the results show that:

- In the first cluster, the lowest amount of error is obtained by the proposed method and lag 3.
- In the second cluster, the lowest amount of error is obtained by the proposed method and lag 1.
- In the third cluster, the lowest amount of error is obtained by the ARIMA method (for all lags)
- In the fourth cluster, the lowest amount of error is obtained by the proposed method and lag 4.
- On average, the lowest amount of error is obtained by the proposed method and lag 6.

Other studies show that the proposed method has a lower prediction error, on average, for all lags, in the dataset of household appliances. This finding means that the proposed model always predicts the behavior of customers better compared to similar designs. Figure 1 shows the mean prediction error of customers' behavior with different plans in the household appliances dataset.

Table 6. Investigating the efficiency of the methods in the dataset of household appliances

Cluster	Method	Lag = 1	Lag = 2	Lag = 3	Lag = 4	Lag = 5	Lag = 6
Cluster 1	SVR	0.1235	0.1034	0.1009	0.0650	0.0608	0.0529
	KNN	0.1396	0.0547	0.0605	0.0368	0.0441	0.0579
	ARIMA	0.0537	0.0537	0.0537	0.0537	0.0537	0.0537
	SVRGOA	0.1344	0.1319	0.1786	0.0638	0.1196	0.1147
	DTWHO	0.1155	0.0638	0.0342	0.0441	0.0572	0.549
Cluster 2	SVR	1.1583	1.1413	1.1504	1.1401	1.1413	1.4059
	KNN	1394	1.3079	1.2500	1.1610	1.1404	1.0772
	ARIMA	1.1819	1.1819	1.1819	1.1819	1.1819	1.1819
	SVRGOA	0.9895	1.1792	1.1865	1.1571	1.5648	1.2551
	DTWHO	0.9832	1.1318	1.1688	1.1475	1.1482	1.094
Cluster 3	SVR	0.1798	0.2089	0.2118	0.1790	0.1653	0.1679
	KNN	0.2943	0.3046	0.2688	0.2745	0.2695	0.3242
	ARIMA	0.1407	0.1407	0.1407	0.1407	0.1407	0.1407
	SVRGOA	0.1817	0.2168	0.1791	0.1956	0.1977	0.2136
	DTWHO	0.1757	0.1932	0.1639	0.1881	0.1798	0.1977
Cluster 4	SVR	0.3455	0.3426	0.3539	0.2742	0.2751	0.4310
	KNN	0.5063	0.4344	0.5307	0.5716	0.5584	0.5231
	ARIMA	0.4078	0.4078	0.4078	0.4078	0.4078	0.4078
	SVRGOA	0.3117	0.4898	0.3301	0.2253	0.3388	0.4752
	DTWHO	0.2876	0.3873	0.3544	0.2169	0.2736	0.2545
Mean	SVR	0.4518	0.4491	0.4542	0.4146	0.4107	0.5144
	KNN	0.5199	0.5254	0.5275	0.5110	0.5031	0.4956
	ARIMA	0.4460	0.4460	0.4460	0.4460	0.4460	0.4460
	SVRGOA	0.4043	0.5044	0.4685	0.4105	0.5552	0.5146
	DTWHO	0.5238	0.4247	0.3991	0.4303	0.4440	0.3905

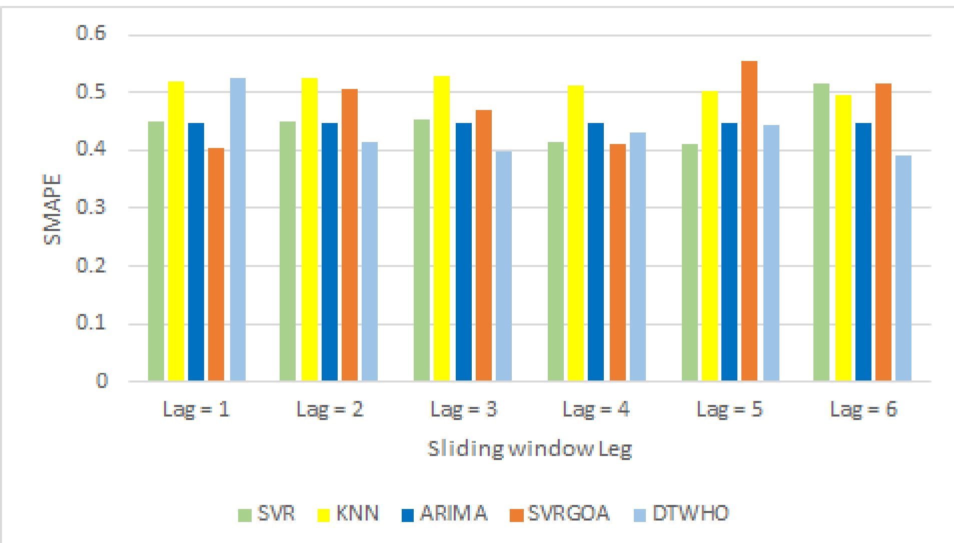

Fig 1. Investigating the mean prediction error of customers' behavior with different plans in the household appliances dataset

b) Predicting the behavior of grocery store customers

In this section, the effectiveness of the proposed model (DTWHO) in predicting customer behavior is investigated. This experiment is performed for different lengths of the time window (1 to 6) and the results are analyzed. It should be mentioned that the results are compared with the similar condition and basic plans, in this review. Table (7) shows the results of this investigation.

Analysis of the results shows that:

- In the first cluster, the lowest amount of error is obtained by the proposed method and lag 3.
- In the second cluster, the lowest amount of error is obtained by the proposed method and lag 5.
- In the third cluster, the lowest amount of error is obtained by the proposed method and lag 2.
- In the fourth cluster, the lowest amount of error is obtained by the ARIMA method (for all lags)
- In the fifth cluster, the lowest amount of error is obtained by the ARIMA method (for all lags)
- On average, the lowest amount of error is obtained by the proposed method and lag 4.

Other investigations show that the mean of the proposed method (for all tests/ the mean of all the results) in the grocery store dataset has a lower prediction error. This finding means that the proposed model predicts customer behavior better than similar plans. Figure 2 shows the mean prediction error of customers' behavior with different plans in the grocery store dataset.

Table 7. Investigating the effectiveness of the methods in the grocery store dataset

Cluster	Method	Lag = 1	Lag = 2	Lag = 3	Lag = 4	Lag = 5	Lag = 6
Cluster 1	SVR	0.2892	0.2861	0.2776	0.2704	0.3014	0.2446
	KNN	0.3018	0.2843	0.2799	0.2540	0.2584	0.2711
	ARIMA	0.3196	0.3196	0.3196	0.3196	0.3196	0.3196
	SVRGOA	0.1972	0.2281	0.1928	0.2267	0.2519	0.2169
	DTWHO	0.1834	0.2198	0.1743	0.2039	0.2367	0.2055
Cluster 2	SVR	0.2210	0.2177	0.2098	0.1348	0.1372	0.1271
	KNN	0.1928	0.1715	0.1135	0.1106	0.0986	0.1043
	ARIMA	0.1679	0.1679	0.1679	0.1679	0.1679	0.1679
	SVRGOA	0.1797	0.1563	0.1847	0.1348	0.2216	0.1596
	DTWHO	0.1794	0.1715	0.1058	0.1065	0.08952	0.0973
Cluster 3	SVR	0.3407	0.3819	0.4066	0.4804	0.5195	0.5099
	KNN	0.3929	0.3682	0.4120	0.3825	0.3962	0.4048
	ARIMA	0.3731	0.3731	0.3731	0.3731	0.3731	0.3731
	SVRGOA	0.3351	0.3713	0.4006	0.5037	0.4804	0.3741
	DTWHO	0.3317	0.3562	0.3980	0.4768	0.4539	0.3675
Cluster 4	SVR	0.0834	0.0904	0.0980	0.0875	0.0686	0.0733
	KNN	0.1275	0.1171	0.1208	0.0801	0.0503	0.0512
	ARIMA	0.0469	0.0469	0.0469	0.0469	0.0469	0.0469
	SVRGOA	0.1125	0.0873	0.0778	0.0968	0.1100	0.0792
	DTWHO	0.0938	0.0856	0.0899	0.0957	0.0785	0.0854
Cluster 5	SVR	0.1253	0.1243	0.1229	0.1380	0.1444	0.1694
	KNN	0.1568	0.1521	0.1111	0.1932	0.1536	0.1672
	ARIMA	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909
	SVRGOA	0.1510	0.1051	0.2506	0.1483	0.2135	0.1753
	DTWHO	0.1187	0.1205	0.1197	0.1191	0.1254	0.1749
Mean	SVR	0.2119	0.2201	0.2230	0.2222	0.2342	0.2249
	KNN	0.2344	0.2186	0.2075	0.2041	0.1914	0.1997
	ARIMA	0.1997	0.1997	0.1997	0.1997	0.1997	0.1997
	SVRGOA	0.1951	0.1896	0.2213	0.2221	0.2555	0.2010
	DTWHO	0.1861	0.1968	0.2004	0.1775	0.1997	0.1814

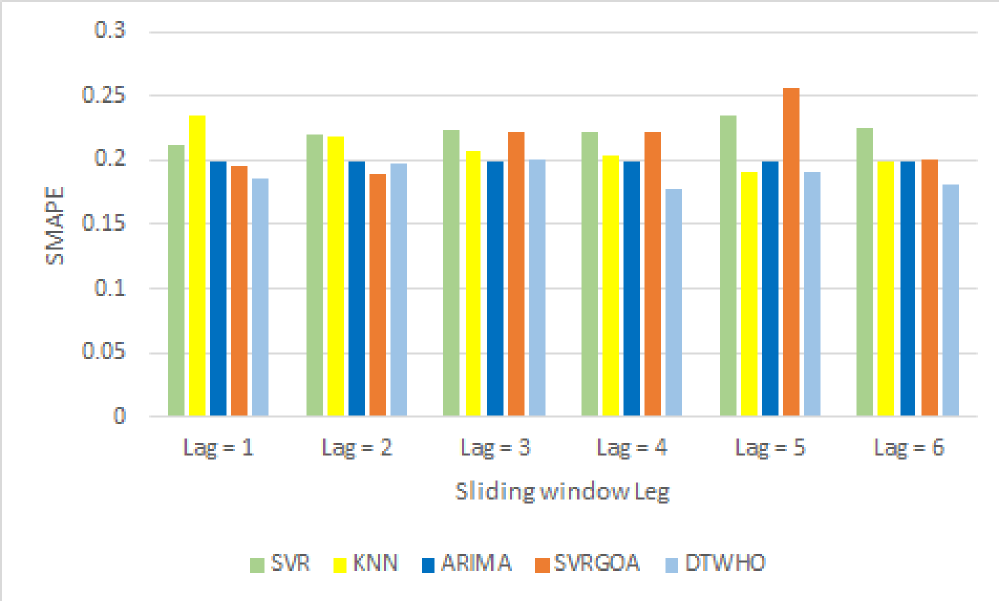

Fig. 2. Investigating the mean prediction error of customers' behavior with different plans in the grocery store dataset

c) Predicting the behavior of supermarket customers

In this section, the effectiveness of the proposed model (DTWHO) in predicting customers' behavior is Investigated.

Table 8. checking the effectiveness of the methods in the data supermarket complex

Cluster	Method	Lag = 1	Lag = 2	Lag = 3	Lag = 4	Lag = 5	Lag = 6
Cluster 1	SVR	0.1107	0.0986	0.0807	0.0466	0.0536	0.0558
	KNN	0.1384	0.0468	0.0516	0.0638	0.0535	0.0589
	ARIMA	0.0336	0.0336	0.0336	0.0336	0.0336	0.0336
	SVRGOA	0.0994	0.1096	0.1224	0.0324	0.1073	0.1278
	DTWHO	0.0976	0.0992	0.1208	0.0.315	0.0994	0.1158
Cluster 2	SVR	0.1642	0.1502	0.1241	0.0641	0.0554	0.0533
	KNN	0.2674	0.1323	0.1232	0.1257	0.0327	0.0307
	ARIMA	0.0284	0.0284	0.0284	0.0284	0.0284	0.0284
	SVRGOA	0.1576	0.1405	0.1613	0.0430	0.1651	0.1431
	DTWHO	0.1472	0.1395	0.1593	0.0455	0.1601	0.1397
Cluster 3	SVR	0.2697	0.2702	0.2687	0.2839	0.2842	0.2820
	KNN	0.2989	0.2078	0.1006	0.1908	0.2498	0.2341
	ARIMA	0.2037	0.2037	0.2037	0.2037	0.2037	0.2037
	SVRGOA	0.3107	0.2554	0.2904	0.2715	0.3398	0.2902
	DTWHO	0.2563	0.1954	0.0998	0.1976	0.2145	0.2066
Cluster 4	SVR	0.1301	0.1067	0.1069	0.0581	0.0551	0.0493
	KNN	0.1908	0.0573	0.0539	0.0455	0.0455	0.0455
	ARIMA	0.0563	0.0563	0.0563	0.0563	0.0563	0.0563
	SVRGOA	0.1492	0.1123	0.1450	0.0510	0.1239	0.1008
	DTWHO	0.1256	0.0843	0.0268	0.0425	0.0508	0.0455
Mean	SVR	0.1687	0.1564	0.1451	0.1132	0.1121	0.1101
	KNN	0.2239	0.1111	0.0823	0.1064	0.0954	0.0923
	ARIMA	0.0805	0.0805	0.0805	0.0805	0.0805	0.0805
	SVRGOA	0.1343	0.1450	0.0564	0.1521	0.1292	0.1463
	DTWHO	0.1269	0.1312	0.0952	0.1181	0.1296	0.1566

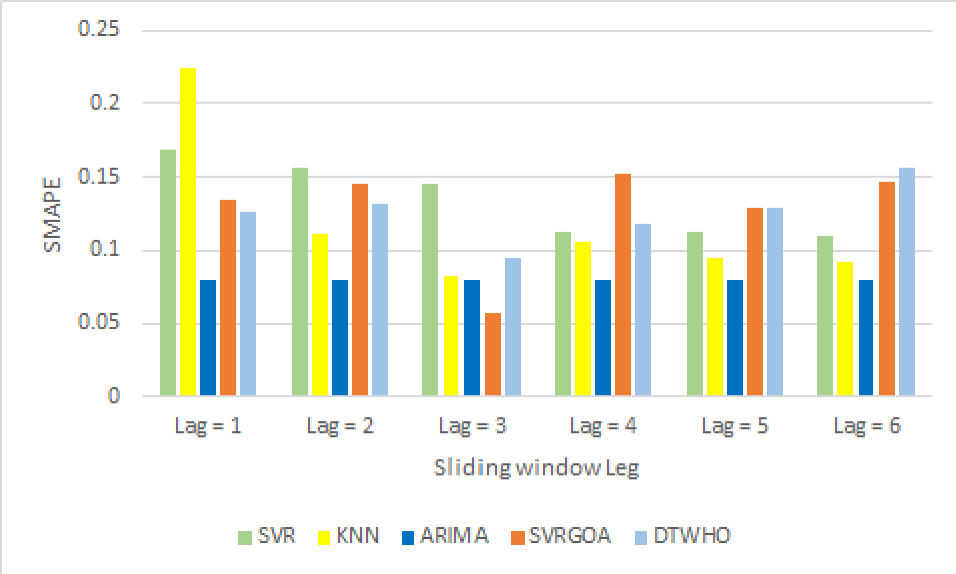

Fig. 3. Mean prediction error of customers' behavior with different designs in the supermarket dataset

This experiment is performed for different lengths of the time window (1 to 6) and the results are analyzed. It should be mentioned that the results are compared with the similar basic plans under the same conditions in this investigation. Table (8) shows the results of this investigation.

Analysis of the results shows that:

- In the first cluster, the lowest amount of error is obtained by the proposed method and lag 4.
- In the second cluster, the lowest amount of error is obtained by the ARIMA method (for all lags)
- In the third cluster, the lowest amount of error is obtained by the proposed method and lag 3.
- In the fourth cluster, the lowest amount of error is obtained by the ARIMA method (for all lags)
- On average, the lowest amount of error is obtained by the ARIMA method (for all lags)

Other reviews show that although the ARIMA model has a lower mean prediction error than the proposed plan; this is the proposed plan in clusters 1 and 3 that predicts customers' behavior by less error. Fig. 3 shows the mean prediction error of customers' behavior with different designs in the supermarket dataset.

6. DISCUSSION AND CONCLUSION

One of the areas in which businesses use artificial intelligence techniques is the analysis and prediction of customer behavior. It is important for a business to anticipate the future behavior of its customers in order to set preventive measures to respond to threats and opportunities in an appropriate manner. In the proposed plan, customer behavior is modeled in the form of time series. The proposed plan includes two general steps.

In the first step, the customers are divided into clusters based on the features extracted from the time series, and in the second step, the customers' behavior is estimated based on an efficient predictive algorithm. To do this, the datasets are first converted into a standard format with the pre-processing process. Then, the customers are separated into different clusters based on the characteristics extracted from the time series. After the completion of the first step, the results are sent to the second step and the behavior of customers is predicted using the improved decision tree algorithm with the wild horse optimization algorithm.

The proposed method of the article was implemented in the MATLAB software environment and its efficiency was evaluated in the Symmetric Mean Absolute Percentage Error (SMAPE) index. For this purpose, the proposed model was implemented along with similar models on three datasets of groceries, household appliances and supermarkets. The results show that 4 features of Var, Spikiness, Lumpiness and Entropy have a high impact intensity among extracted features. Other results also show that the best number of clusters for clustering customers in the grocery dataset is 5, for the household appliances dataset is 4, and for the supermarket dataset is 4. After the clustering of customers, the process of predicting customers' behavior starts and the results show that the lowest prediction error is achieved by the proposed method and lag 6 in the dataset of household appliances. In the grocery store dataset, the lowest prediction error is obtained by the proposed method and lag 4, but in the supermarket dataset, the ARIMA-based model has a lower mean prediction error, but it should be kept in mind that the proposed plan predicts customer behavior with less error for clusters 1 and 3.

REFERENCES

[1] P. Li, T. Jiang, L. Liu, and S. I. Ammar, **"Customer Behavior Analysis in Ticket Windows at Scenic Spots from Game-Theoretic Perspective,"** *Discrete Dynamics in Nature and Society*, vol. 2021, 2021.

[2] M. Liu, L. Zhao, F. Sun, W. Zhao, Y. Zuo, and K. Yada, **"Customer Behavior Analysis and Classification Based on Process Mining,"** *in 2021 IEEE International Conference on Systems*, Man, and Cybernetics (SMC), 2021, pp. 1000-1005: IEEE.

[3] N. T. Hien, Y.-L. Su, R. Sann, and L. T. P. Thanh, **"Analysis of Online Customer Complaint Behavior in Vietnam's Hotel Industry,"** *Sustainability*, vol. 14, no. 7, p. 3770, 2022.

[4] R. A. Daoud, A. Amine, B. Bouikhalene, and R. Lbibb, **"Combining RFM model and clustering techniques for customer value analysis of a company selling online,"** *in 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, 2015, pp. 1-6: IEEE.

[5] G. Gartner, **"CRM Economics: Figuring Out the ROI on Customer Initiative,"** White Paper, Stamford, CT, 2001.

[6] Y. M. Baashar, A. K. Mahomood, M. A. Almomani, and G. A. Alkawsi, **"Customer relationship management (CRM) in healthcare organization: A review of ten years of research,"** *in Computer and Information Sciences (ICCOINS), 2016 3rd International Conference on*, 2016, pp. 97-102: IEEE.

[7] Y. Saberi, M. Ramezanpour, and R. Khorsand, **"An efficient data hiding method using the intra prediction modes in HEVC,"** *Multimedia Tools and Applications* 79 (2020): 33279-33302.

[8] I. Dalla Pozza, O. Goetz, and J. M. Sahut, **"Implementation effects in the relationship between CRM and its performance,"** *Journal of Business Research*, vol. 89, pp. 391-403, 2018.

[9] A. D. Stein, M. F. Smith, and R. A. Lancioni**, "The development and diffusion of customer relationship management (CRM) intelligence in business-to-business environments,"** *Industrial Marketing Management*, vol. 42, no. 6, pp. 855-861, 2013.

[10] S. K. I. Shaon and H. Rahman, **"A theoretical review of CRM effects on customer satisfaction and loyalty,"** *Central European Business Review*, vol. 4, no. 1, p. 23, 2015.

[11] A. Josiassen, A. G. Assaf, and L. K. Cvelbar, **"CRM and the bottom line: Do all CRM dimensions affect firm performance?,"** *International Journal of Hospitality Management*, vol. 36, pp. 130-136, 2014.

[12] N. Deluxni, **"Comparative Analysis of Customer Behaviour Using K-Means Algorithm Over Convolutional Neural Network with Increase Inaccuracy of Prediction,"** *ECS Transactions*, vol. 107, no. 1, p. 12459, 2022.

[13] L. M. Aouad, N.-A. Le-Khac, and T. Kechadi, **"Variance-based clustering technique for distributed data mining applications,"** arXiv preprint arXiv:1703.09823, 2017.

[14] Y. T. Tang, H.-H. Wu, H.-W. Yang, J. I. Shieh, and M. M. Lo, **"Customer Behavior Analysis of the Chinese New Year from a Supermarket in Taiwan,"** *in 2022 2nd International Conference on Information Technology and Education (ICIT&E)*, 2022, pp. 156-159: IEEE.

[15] M. Nazer, and R. Khorsand, **"Energy Aware Resource Provisioning for Multi-Criteria Scheduling in Cloud Computing,"** *Cybernetics and Systems*, 2022 Apr 30:1-30.

[16] H. Abbasimehr and M. Shabani, **"A new methodology for customer behavior analysis using time series clustering,"** *Kybernetes*, 2019.

[17] I. Naruei and F. Keynia, **"Wild horse optimizer: A new meta-heuristic algorithm for solving engineering optimization problems,"** *Engineering with Computers*, pp. 1-32, 2021.

[18] H. Abbasimehr and M. Shabani, **"A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers,"** Kybernetes, 2019.

[19] W. Chang, C. Chang, and Q. Li, "**Customer lifetime value: A review,"** *Social Behavior and Personality: an international journa*l, vol. 40, no. 7, pp. 1057-1064, 2012.

[20] R. J. Freund, W. J. Wilson, and P. Sa, Regression analysis. Elsevier, 2006.

[21] A. O. Sykes, **"An introduction to regression analysis,"** 1993.

[22] L. Gordon, **"Using Classification and Regression Trees (CART) in SAS® Enterprise Miner TM For Applications in Public Health,"** *Public Health*, 2013.

[23] H. Abbasimehr and F. S. Baghery, **"A novel time series clustering method with fine-tuned support vector regression for customer behavior analysis,"** E*xpert Systems with Applications*, p. 117584, 2022.

[24] R. Siva Subramanian and D. Prabha, **"Customer behavior analysis using Naive Bayes with bagging homogeneous feature selection approach,"** *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 5105-5116, 2021.

[25] A. C. Gopal and L. Jacob, **"Customer Behavior Analysis Using Unsupervised Clustering and Profiling: A Machine Learning Approach,"** i*n 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE),* 2022, pp. 2075-2078: IEEE.

[26] M. Hosseini, N. Abdolvand, and S. R. Harandi, **"Two-dimensional analysis of customer b**ehavior i**n traditional and electronic banking**," *Digital Business*, vol. 2, no. 2, p. 100030, 2022.

[27] O. Dogan, O. F. Seymen, and A. Hiziroglu, **"Customer Behavior Analysis by Intuitionistic Fuzzy Segmentation: Comparison of Two Major Cities in Turkey,"** I*nternational Journal of Information Technology & Decision Making*, vol. 21, no. 02, pp. 707-727, 2022.

[28] T. Hlupić, D. Oreščanin, and M. Baranović, **"A Novel Method for IPTV Customer Behavior Analysis Using Time Series,"** *IEEE Access*, vol. 10, pp. 37003-37015, 2022.

[29] V. Sundararaj and M. Rejeesh, **"A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites,"** *Journal of Retailing and Consumer Services*, vol. 58, p. 102190, 2021.

[30] L. Ma, Z. Li, and D. Zheng, **"Analysis of Chinese consumers' willingness and behavioral change to purchase Green agri-food product online,"** *Plos one*, vol. 17, no. 4, p. e0265887, 2022.

[31] F. Baratzadeh and S. M. Hasheminejad, **"Customer Behavior Analysis to Improve Detection of Fraudulent Transactions using Deep Learning,"** *Journal of AI and Data Mining*, vol. 10, no. 1, pp. 87-101, 2022.

[32] H. Abbasimehr and M. Shabani, **"A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series techniques,"** *Journal of ambient intelligence and humanized computing*, vol. 12, no. 1, pp. 515-531, 2021.

[33] K. Bandara, C. Bergmeir, and S. Smyl, **"Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach,"** *Expert systems with applications*, vol. 140, p. 112896, 2020.

[34] R. J. Hyndman, E. Wang, and N. Laptev, **"Large-scale unusual time series detection," i**n *2015 IEEE international conference on data mining workshop (ICDMW)*, 2015, pp. 1616-1619: IEEE.

[35] X. He, D. Cai, and P. Niyogi, **"Laplacian score for feature selection,"** *Advances in neural information processing systems*, vol. 18, pp. 507-514, 2005.

[36] T. M. Silva Filho, B. A. Pimentel, R. M. C. R. Souza, and A. L. I. Oliveira, **"Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization,"** *Expert Systems with Applications*, vol. 42, no. 17, pp. 6315-6328, 2015/10/01/ 2015.

[37] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, **"Text classification algorithms: A survey,"** *Information*, vol. 10, no. 4, p. 150, 2019.

[38] J. Nayak, B. Naik, and H. Behera, **"A comprehensive survey on support vector machine in data mining tasks: applications & challenges,"** *International Journal of Database Theory and Application*, vol. 8, no. 1, pp. 169-186, 2015.

[39] X. Zhang and Y. Guo, **"Optimization of SVM parameters based on PSO algorithm,"** *in 2009 Fifth International Conference on Natural Computation*, 2009, vol. 1, pp. 536-539: IEEE.

[40] M. Mafarja, I. Aljarah, H. Faris, A. I. Hammouri, A.-Z. Ala'M, and S. Mirjalili, **"Binary grasshopper optimisation algorithm approaches for feature selection problems,"** *Expert Systems with Applications*, vol. 117, pp. 267-286, 2019.

[41] F. Martínez, M. P. Frías, M. D. Pérez-Godoy, and A. J. Rivera, **"Dealing with seasonality by narrowing the training set in time series forecasting with kNN,"** *Expert systems with applications*, vol. 103, pp. 38-48, 2018.