An Improved Decision Tree Classification Method based on 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: 9 March 2023 Revised: 24 April 2023 Accepted: 14 May 2023

ABSTRACT:
In this paper, an improved decision tree classification method based on wild horse optimization algorithm is proposed and then the application in customer behavior analysis is evaluated. Customer behavior is modeled in the form of time series. The proposed method includes two general steps. First, the customers are classified into clusters based on the features extracted from the time series, and then the customers' behavior is estimated based on an efficient predictive algorithm in the second step. In this paper, an improved decision tree classification based on wild horse optimization algorithm is used to predict customer behavior. The proposed method is implemented in the MATLAB software environment and its efficiency is evaluated in the Symmetric Mean Absolute Percentage Error (SMAPE) index. The experimental results show that variance, spikiness, lumpiness and entropy have a high impact intensity among the extracted features. The overall evaluation indicate that this proposed method obtains the lowest prediction error in compared to other evaluated methods.

KEYWORDS: Customer behavior, Decision tree, Classification, Prediction.

1. INTRODUCTION

Companies must predict the market situation with the increase in competition (business fields) from raw data and information analysis in the next few periods and also predict the behavior of customers for the future in order to maintain their survival in today's era. [1]. Because the customer is considered as the main key for the success or failure of a company [2]. One of the most important areas of study in this field is the investigation of customer loyalty, satisfaction, preferences and taste, which researchers try to analyze customer behavior by using different methods and tools [3].

Customer relationship management is a model that has provided the basis for the analysis of raw customer data [4]. Customer relationship management is a popular and strategic topic in marketing and service quality [4]. Access to extensive transactional data as well as computing systems have provided the best opportunity for modeling and predicting customer behavior [5].

Customer clustering is one of the main steps to analyze customer behavior in the process of customer relationship management [6]. Cluster analysis is a method for grouping data or observations according to their similarity and degree of closeness, and data or observations are divided into homogeneous or distinct categories through cluster analysis [7]. A customer's behavior can be dynamic due to various factors; it means that the customer behaves differently in different periods of time. Accordingly, efficient clustering is a model that groups customers based on time series features. The behavior analysis model will be a dynamic model based on such an approach in clustering. It is necessary for such a process to extract the features of the time series in the first step; then, the clustering process is performed. The current article also focuses on clustering based on time series features this approach due to its efficiency [8].

The time series of the samples is called and their means are calculated for each cluster after clustering the samples in the proposed scheme. The resulting mean is considered as the representative of the cluster. Customer behavior can be estimated by relying on representative vectors and using an efficient predictive algorithm. The proposed algorithm of the article is for prediction of the decision tree [9].

Decision tree is a decision support tool that uses trees for modeling [10]. Decision trees are commonly used in various machine learning applications. The application of decision tree is in the description of conditional probability calculations. Decision tree is a supervised

Paper type: Research paper
DOI: 10.30486/mjtd.2023.1980072.1026
How to cite this paper: R. Sharifi, M. Ramezanpour, "**An Improved Decision Tree Classification Method based on Wild Horse Optimization Algorithm**", *Majlesi Journal of Telecommunication Devices*, Vol. 12, No. 4, pp.177-188, 2023.

learning algorithm and it is used for data classification and 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 [11]. The leaves in the tree structure represent the class labels and the branches represent the relation of the features that are branched to these labels.

But it should be kept in mind that every predictive algorithm has parameters to adjust; it can be presented in the form of an optimization problem what are the values of these parameters; so that the algorithm predicts the forecasting process with the lowest amount of error [12]. Several optimization algorithms have been proposed in recent years, but one of the algorithms that is less trapped in the local optimum and provides better results in comparison with the similar algorithms is the wild horse optimization algorithm [13]. 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.

The wild horse optimization algorithm is considered an efficient meta-heuristic algorithm that has been used for single-purpose and multi-purpose problems. Its binary version is also available for discrete problems. Fig. 1 shows the structure of the initial population in the optimization algorithm of wild horses.

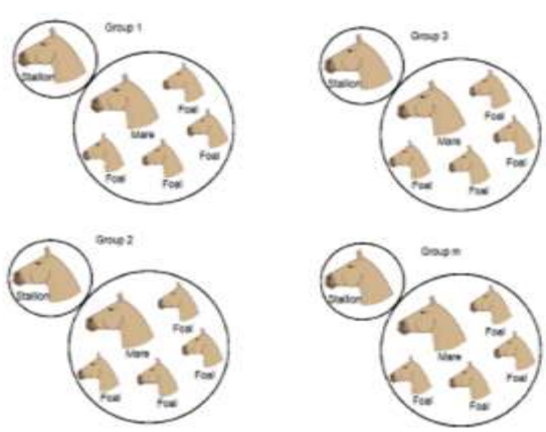

Fig. 1. A view of the production of the initial population of wild horses [13].

The most important purpose of the article is to analyze the behavior of household appliance customers which will be done by clustering based on time series features and improved decision tree with wild horse optimization algorithm. The proposed approach will try to reduce the behavior prediction error while properly clustering household appliance customers.

2. RELATED WORKS

A new method for analyzing customer behavior is presented in [14] by combining clustering algorithms and time series forecasting. Time series features (13 frequently used features)are extracted for each customer in the clustering step, after the pre-processing operation on the data. Then, four effective features are selected for the clustering process using the Laplace scoring method, and the samples are clustered based on the effective features. Four features of the variance of all points of the time series (variance), spectral entropy (entropy), change of variance in the residual (lumpiness) and the amount of stickiness of the time series (spikiness) have the highest score among the characteristics in this step. The mean of the clusters is calculated after clustering and the obtained vector is considered as the representative of the cluster. Improved support vector regression with grasshopper optimization algorithm (SVRGOA) and ARIMA, SVR and KNN algorithms are used to estimate customer behavior in the second step of the proposed method. Support vector regression is combined with meta-heuristic algorithm in order to increase efficiency and optimal setting of parameters. Examining the results showed that the use of feature-based clustering has better results than distance-based clustering in the analysis of customer behavior.

In [15], instead of clustering on the time series itself, first a number of features were extracted from the time series based on the features of [16]; 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.

In [7], 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 [17] 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.

The approach of the article [18] is to combine the time factor in customer segmentation using RFM which includes customer segmentation using K-means algorithm in each time period and labeling each segment then tracking customer behavior over time.

Researchers have used binary logistic regression in [19] to investigate the buying behavior of consumers of agricultural products. The result of the research shows that gender, age, consumers’ understanding of the product and monthly income have a positive effect on the stability of consumers’ behavior and willingness to buy; While the price of agricultural products and online shopping have a negative effect on the willingness of consumers.

The article [20] 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.

The approach of the article [18] is to combine the time factor in customer segmentation using RFM which includes customer segmentation using K-means algorithm in each time period and labeling each segment and then tracking customer behavior over time.

3. 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 method 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 method includes two general steps. The customers are seperated 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. The representative of each cluster is extracted in the following. Fig (2) shows the general diagram of the first phase of the proposed plan.

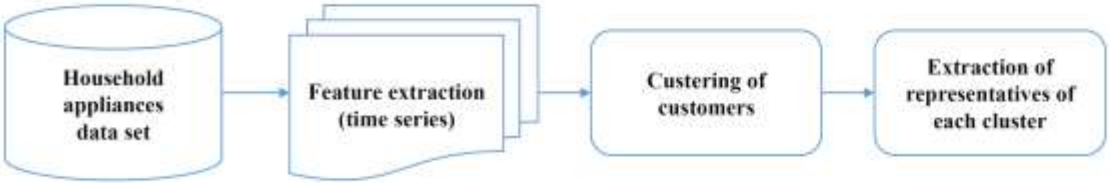

Fig. 2. Diagram of the first phase of the proposed method.

The results are sent to the second step after the end of the first step and the behavior of customers is predicted using the improved decision tree algorithm with the wild horse optimization algorithm.

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 method 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 [13] can be a suitable tool for this purpose.

The improved decision tree with wild horse perform the customer behavior prediction process on the representative vectors of the clusters and predicts the customers' behavior. And the prediction error is calculated for each cluster with each of the methods by using a known error evaluation criterion. Fig (3) shows the diagram of the second step of the proposed plan.

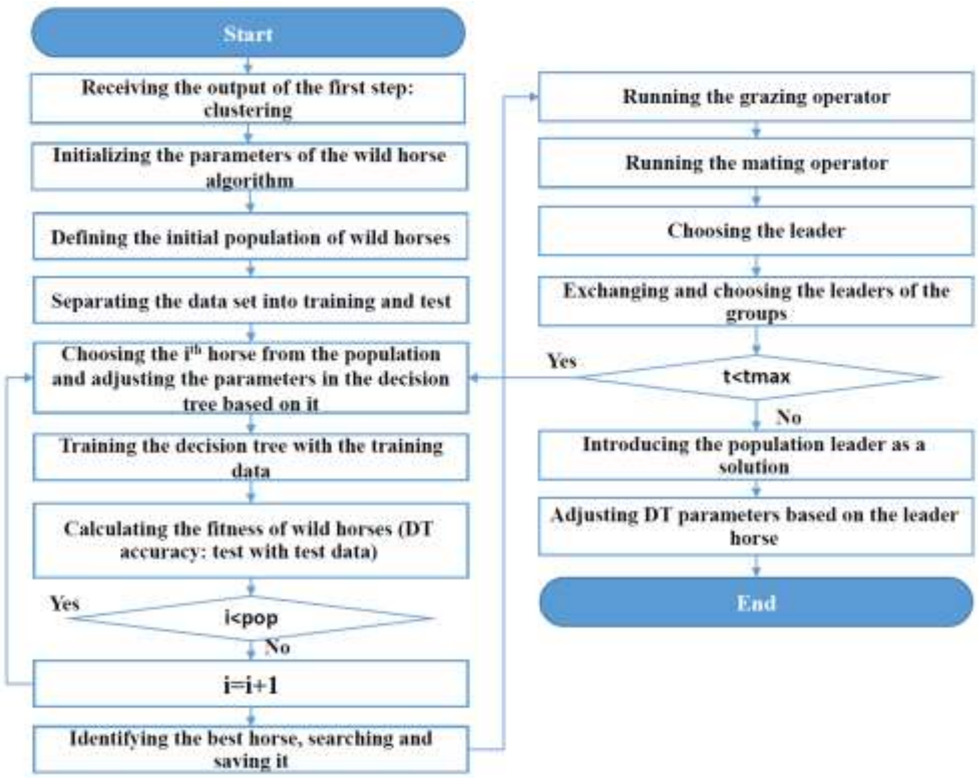

Fig 3. Diagram of the second step of the proposed method.

As it mentioned the proposed method includes two general steps of clustering and prediction; each of these steps is briefly described in the following section.

3.1. 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 method is performed based on the approach of article [15]. Table (1) introduces the list of features that are extracted in this step.

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 [21]. 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. After clustering the samples (based on the valuable features) in the proposed scheme, 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.

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

3.2. 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 [22]. 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 [23]. 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 [24]. Since the use of meta-heuristic optimization algorithms is considered a suitable approach to solve NP-Hard problems [25]; 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 scheme 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. In the proposed plan, because it is intended to adjust 6 parameters of the decision tree; therefore, the length vector of each wild horse will be as 6.

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 $G = [N \times PS]$. 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 function in the proposed method is reduction of the prediction error rate by the decision tree, which is modeled according to Eq. (1). 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.

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

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 method is the modeling of grazing behavior. 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. (2) 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).

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

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. (3), 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.

$$P = \vec{R}_1 < \text{TDR}; \quad IDX = (P == 0); \quad Z = R_2 \odot \text{IDX} + \vec{R}_3 \odot (\sim \text{IDX}) \quad (3)$$

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. (4); this parameter reaches 0 at the end of the algorithm run.

$$\text{TDR} = 1 - \text{iter} \times \left(\frac{1}{\text{maxiter}}\right) \quad (4)$$

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 method is the mating operator. 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. (5) is proposed to simulate the behavior of exit and the mating of horses. This equation is the mean-type crossover operator.

$$X_{G,K}^{p} = \text{Crossover}(X_{G,i}^{q}, X_{G,j}^{z}) \quad i \neq j \neq k, p = q = \text{end} \quad (5)$$
$$\text{Crossover} = \text{Mean}$$

where $X_{G,K}^{P}$ 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, $X_{G,i}^{q}$ 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 $X_{G,j}^{Z}$ 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. (6) is proposed to perform this operator in the wild horse optimization algorithm.

$$\overline{\mathrm{Stallion}_{G_i}} = \begin{Bmatrix} 2Z\cos(2\pi RZ) \times (\mathrm{WH} - \mathrm{Stallion}_{G_i}) + \mathrm{WH} & \text{if } R_3 > 0.5 \\ 2Z\cos(2\pi RZ) \times (\mathrm{WH} - \mathrm{Stallion}_{G_i}) - \mathrm{WH} & \text{if } R_3 \leq 0.5 \end{Bmatrix} \quad (6)$$

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. (7).

$$\mathrm{Stallion}_{G_i} = \begin{Bmatrix} X_{G,i} & \text{if } \cos t\left(X_{G,i}\right) < \cos t(\mathrm{Stallion}_{G_i}) \\ \mathrm{Stallion}_{G_i} & \text{if } \cos t\left(X_{G,i}\right) > \cos t(\mathrm{Stallion}_{G_i}) \end{Bmatrix} \quad (7)$$

4. 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. (8) formulates this index [26].

$$SMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|\hat{y_t} - y_t|}{\frac{|\hat{y_t}| + |y_t|}{2}} \quad (8)$$

In (8), 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 the dataset of household appliances. The research dataset consists of 11-month POS device data collected from different stores. The total number of terminals is 416 which have regular transactions during 44 weeks. Table (3) introduces the generalities related to these datasets.

Table 3. Introducing the datasets used in the research

Dataset title	Number of terminals	Number of transactions	Number of clusters
Household appliances	416	44 series per terminal	4

4.1. Selection of effective features

The clustering process is done in the proposed method 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. Fig. (4) shows the results of this survey.

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.

4.2. 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 [14], 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 fig. (5) indicates that the best case for the household appliance dataset is with 4 clusters. The results of [14] are documented in this article.

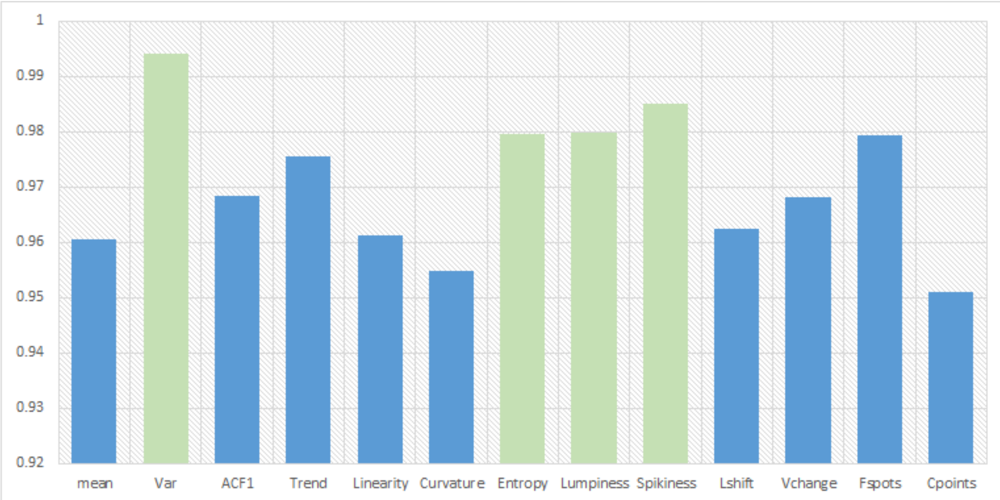

Fig 4. Feature score based on the Laplacian approach.

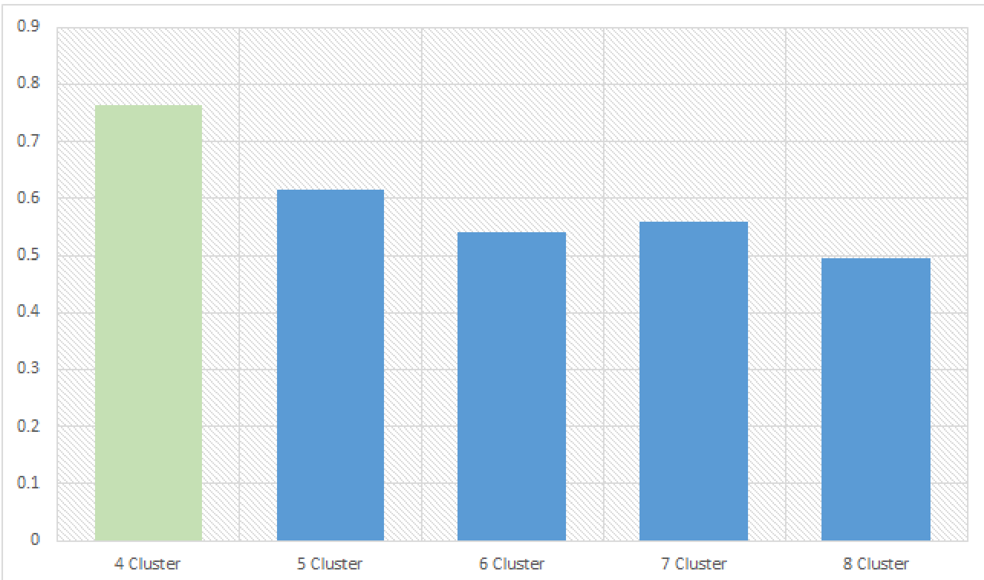

Fig. 5. Customer clustering results in different datasets.

4.3. Predicting customer behavior

The performance of the proposed method is evaluated on the household appliances dataset by the SMAPE index, and the results is compared with similar works under similar conditions. The results of the first step in the proposed method (clustering) show that the number of representative vectors of the clusters (in the best possible case) is 4 for the household appliances dataset. Each representative vector is converted into a dataset in this step using the sliding window and is used for prediction. The length of the time window is considered as a control parameter in this experiment 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 is achieved by changing it. For this purpose, tests are performed for different lengths of the time window (1 to 6).

Examining the results for the first cluster shows that the lowest amount of error is obtained by the proposed method in this cluster and lag 3 is achieved. Fig. 6 shows the results of this investigation.

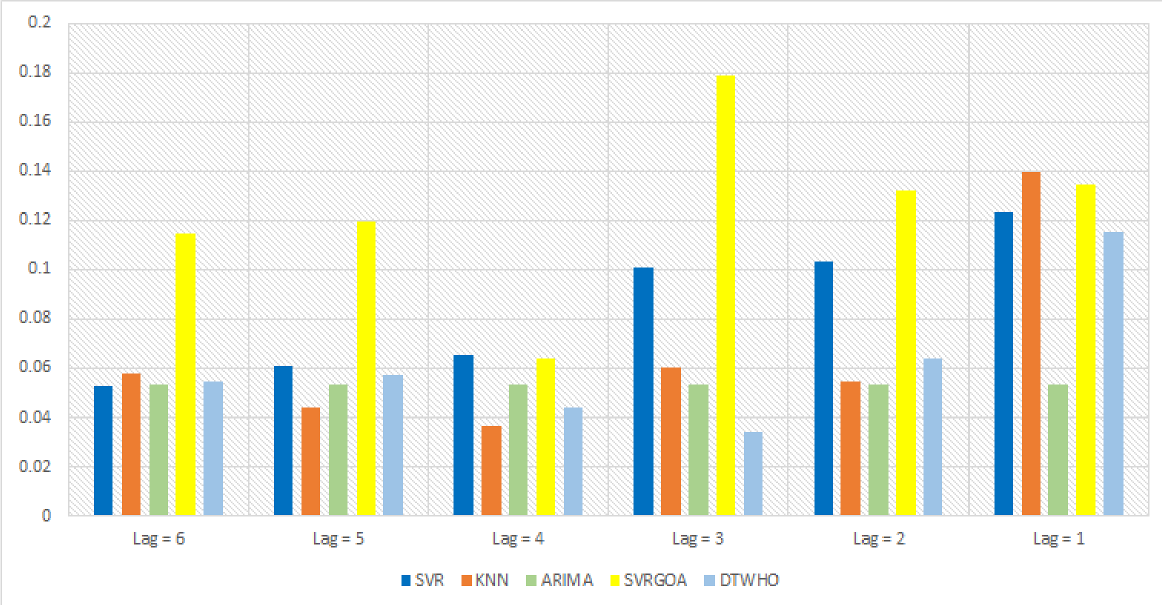

Fig. 6. Prediction error of customer behavior in the first cluster with different methods.

Other investigations for the second cluster show that the lowest amount of error is obtained in this cluster with the proposed method and lag 1 is achieved. Fig. 7 shows the results of this investigation.

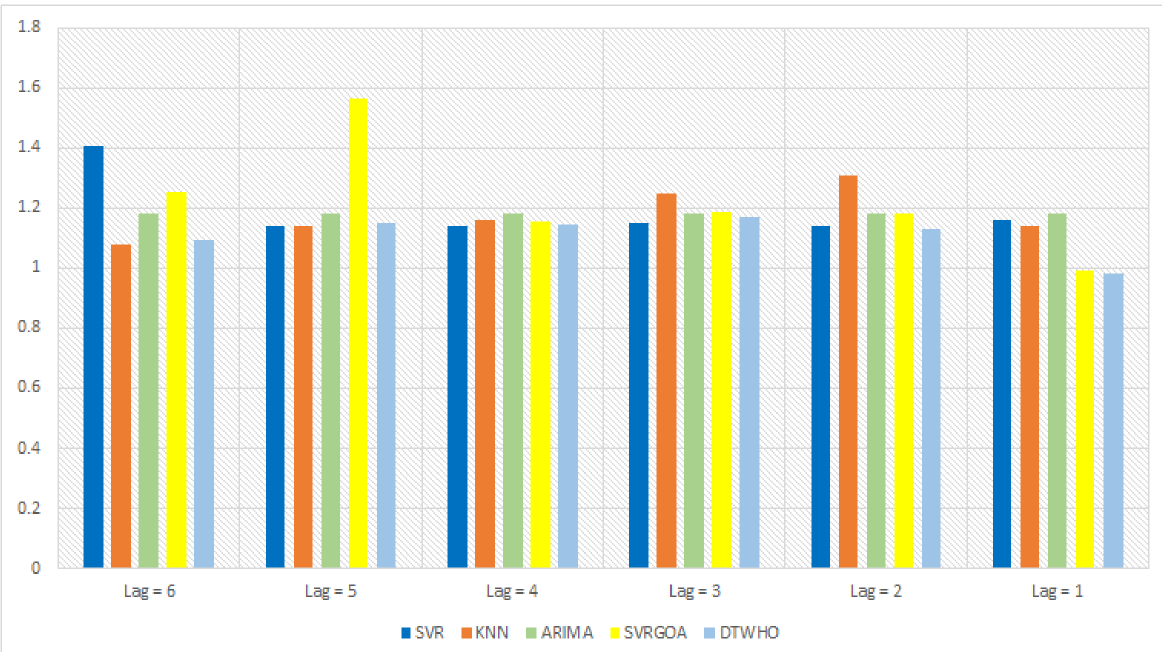

Fig. 7. Prediction error of customer behavior in the second cluster with different methods.

The reviews for the third cluster also indicate that the lowest amount of error in this cluster is obtained by the ARIMA method (for all lags). Fig. 8 shows the results of this investigation.

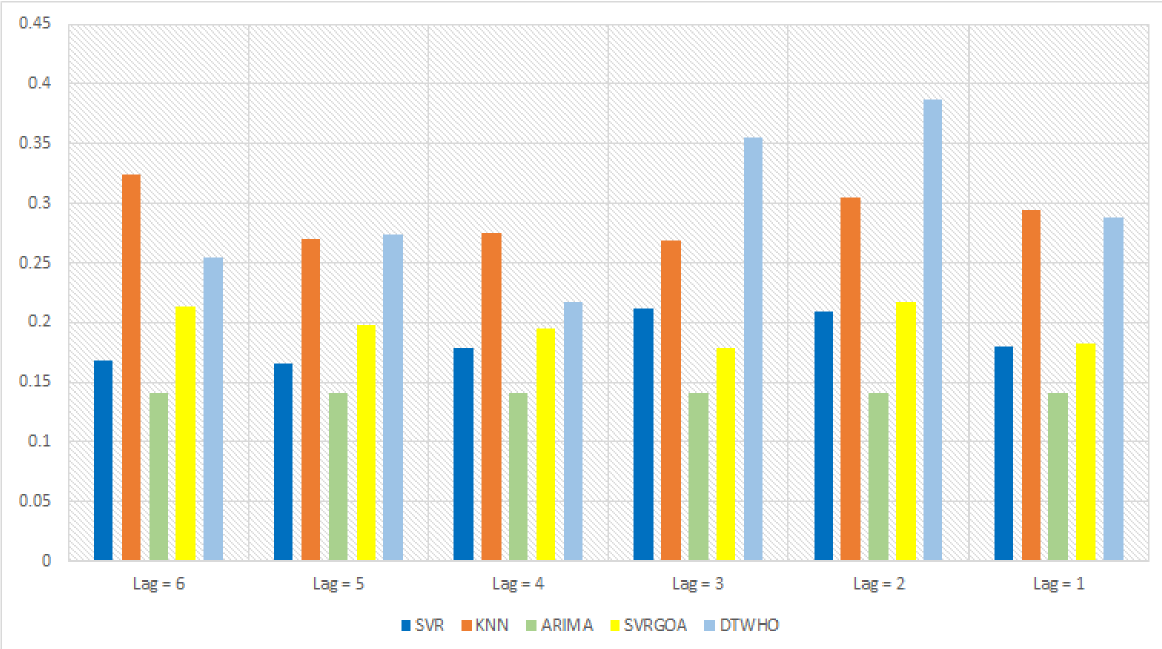

Fig. 8. Prediction error of customer behavior in the third cluster with different methods.

Finally, the results of the investigations in the fourth cluster show that the least amount of error in this cluster is obtained by the least amount of error with the proposed method and lag 4. Fig. 9 shows the results of this investigation.

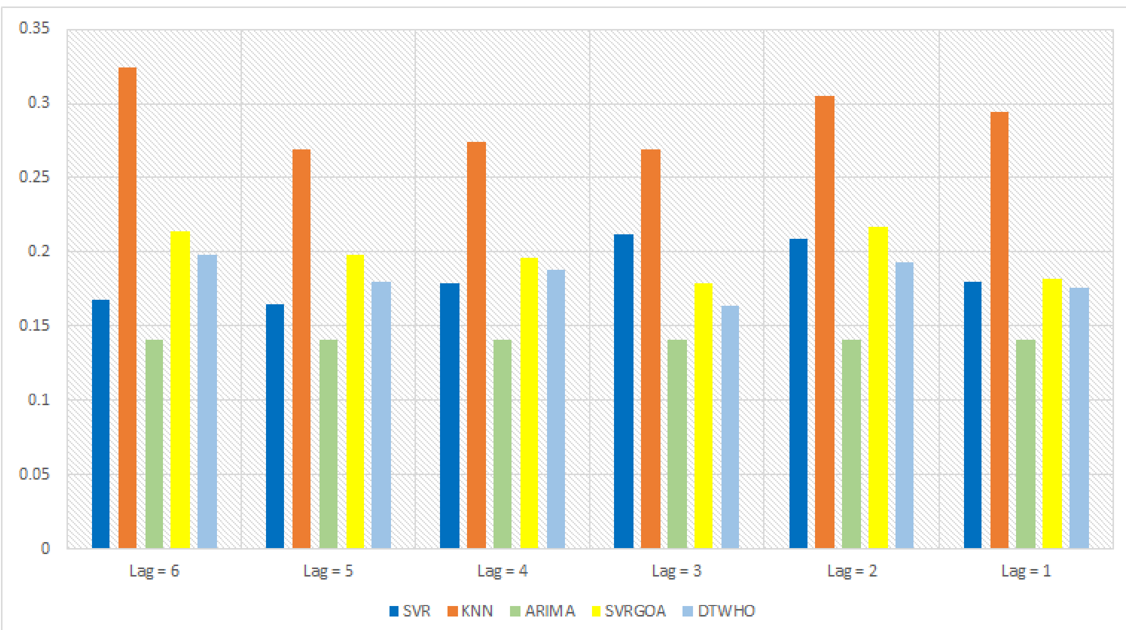

Fig. 9. Prediction error of customer behavior in the fourth cluster with different methods.

The mean prediction of customer behavior (for different clusters) with different algorithms shows that the proposed method has less prediction error in comparison with similar algorithms. So that the lowest amount of error is obtained by the proposed method and lag 6 is achieved. Fig. 10 shows the results of this investigation.

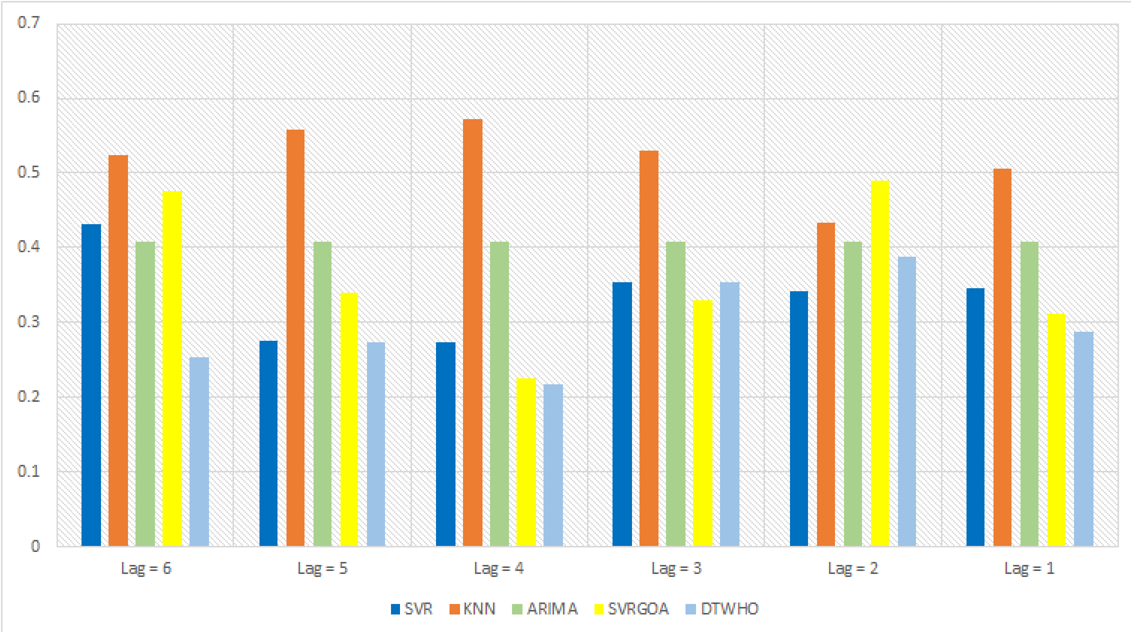

Fig. 10. Mean prediction error of customer behavior for different clusters.

The results of Fig. 10 mean that the proposed model always predicts better customers' behavior when is compared to similar plans.

5. DISCUSSION AND CONCLUSION

Methods based on artificial intelligence and machine learning have become one of the most important forecasting approaches in the decision-making process and macro investments in the markets in the current era. Meanwhile, accuracy of forecasting is one of the most important indicators for choosing a forecasting method and algorithm. Experience is shown that by using machine learning techniques, it is possible to analyze customer behavior and identify hidden knowledge in customer behavior. Predicting customer behavior plays a constructive role in marketing and profitability. It has been tried in this article to provide an efficient model for understanding and predicting customer behavior by combining different artificial intelligence approaches. In the proposed plan, customer behavior is modeled in the form of time series. The proposed method 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 is implemented in the MATLAB software environment and its efficiency is evaluated in the Symmetric Mean Absolute Percentage Error (SMAPE) index. The results showed that 4 features of variance, spikiness, lumpiness and entropy have a high impact intensity among extracted features. Other results also showed that the best number of clusters for clustering customers in the household appliances dataset is 4. After the clustering of customers, the process of predicting customers' behavior has started and the results show that the mean lowest prediction error is achieved by the household appliances dataset by the proposed method and lag 6 is obtained.

REFERENCES

[1] 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.

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

[3] 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.

[4] M. Nazer, and R. Khorsand, **"Energy Aware Resource Provisioning for Multi-Criteria**

Scheduling in Cloud Computing," *Cybernetics and Systems*, 2022 Apr 30:1-30.

[5] 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.

[6] 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.

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

[8] F. Tavousi, S. Azizi, and A. Ghaderzadeh, **"A fuzzy approach for optimal placement of IoT applications in fog-cloud computing,"** *Cluster Computing*, vol. 25, no. 1, pp. 303-320, 2022.

[9] L. Breiman, Classification and regression trees. Routledge, 2017.

[10] 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.

[11] S. B. Kotsiantis, **"Decision trees: a recent overview,"** *Artificial Intelligence Review*, vol. 39, no. 4, pp. 261-283, 2013.

[12] A. Tharwat and A. E. Hassanien, **"Chaotic antlion algorithm for parameter optimization of support vector machine,"** *Applied Intelligence*, vol. 48, no. 3, pp. 670-686, 2018.

[13] R Sharifi, M Ramezanpour, **"Customer Behavior Analysis using Wild Horse Optimization Algorithm,"** *Majlesi Journal of Telecommunication Devices*, 2023.

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

[15] 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.

[16] M. Norouzi, A. Arshaghi, and M. Ashourian, **"Encryption of Color Images using Pixel Shift Algorithm and Developed Hill Algorithm,"** *Majlesi Journal of Telecommunocation Device*s, vol. 11, no. 4, pp.177-184, 2022.

[17] M. Hosseini, N. Abdolvand, and S. R. Harandi, **"Two-dimensional analysis of customer behavior in traditional and electronic banking,"** *Digital Business*, vol. 2, no. 2, p. 100030, 2022.

[18] M. Hosseini and M. Shabani, **"New approach to customer segmentation based on changes in customer value,"** *Journal of Marketing Analytics*, vol. 3, no. 3, pp. 110-121, 2015.

[19] 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.

[20] 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.

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

[22] 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.

[23] 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 Applicat*ion, vol. 8, no. 1, pp. 169-186, 2015.

[24] 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.

[25] 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.

[26] 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.