



Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector

Irina V. Pustokhina¹ · Denis A. Pustokhin² · Phong Thanh Nguyen³ · Mohamed Elhoseny^{4,5} · K. Shankar⁶

Received: 6 September 2020 / Accepted: 25 March 2021
© The Author(s) 2021

Abstract

Customer retention is a major challenge in several business sectors and diverse companies identify the customer churn prediction (CCP) as an important process for retaining the customers. CCP in the telecommunication sector has become an essential need owing to a rise in the number of the telecommunication service providers. Recently, machine learning (ML) and deep learning (DL) models have begun to develop effective CCP model. This paper presents a new improved synthetic minority over-sampling technique (SMOTE) with optimal weighted extreme machine learning (OWELM) called the ISMOTE-OWELM model for CCP. The presented model comprises preprocessing, balancing the unbalanced dataset, and classification. The multi-objective rain optimization algorithm (MOROA) is used for two purposes: determining the optimal sampling rate of SMOTE and parameter tuning of WELM. Initially, the customer data involve data normalization and class labeling. Then, the ISMOTE is employed to handle the imbalanced dataset where the rain optimization algorithm (ROA) is applied to determine the optimal sampling rate. At last, the WELM model is applied to determine the class labels of the applied data. Extensive experimentation is carried out to ensure the ISMOTE-OWELM model against the CCP Telecommunication dataset. The simulation outcome portrayed that the ISMOTE-OWELM model is superior to other models with the accuracy of 0.94, 0.92, 0.909 on the applied dataset I, II, and III, respectively.

Keywords Customer churn · Customer retention · Predictive modeling · Machine learning · Rain optimization algorithm

✉ K. Shankar
drkshankar@ieee.org
Irina V. Pustokhina
ivpustokhina@yandex.ru
Denis A. Pustokhin
dpustokhin@yandex.ru
Phong Thanh Nguyen
phong.nt@ou.edu.vn
Mohamed Elhoseny
melhoseny@ieee.org

- ¹ Department of Entrepreneurship and Logistics, Plekhanov Russian University of Economics, 117997 Moscow, Russia
- ² Department of Logistics, State University of Management, 109542 Moscow, Russia
- ³ Department of Project Management, Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam
- ⁴ College of Computer Information Technology, American University in the Emirates, Dubai, United Arab Emirates
- ⁵ Faculty of Computers and Information, Mansoura University, Mansoura, Egypt
- ⁶ Department of Computer Applications, Alagappa University, Karaikudi, India

Introduction

Customer retention plays an important role in churn prediction, and telecommunication organization is directly associated with financial companies. Hence, many industries performed different actions for developing a robust correlation with users and to reduce user defections. In order to establish the best customer retention, it is essential to understand the user changing hierarchy and find the causes involved in churning. In recent times, the competing nature and cost-cutting stress invoked the firms to maximize the customer relationship management (CRM) module. The consequent nature of the customer has to be changed from unknown to known ones. But the prediction of the user's decision is highly complex and early detection helps prevent the churn prediction [1]. It is a mandatory input for customer retention models and extraordinary events like gifts, incentives, discounts, and promotions would improve the user requirement and satisfaction. The impact of customer satisfaction on retention in the telecommunication industry is essential, and expert's knowledge regarding the related aspects

influences the user satisfaction to enhance limit the churn, where CRM should be employed proficiently [2].

In CRM, user maintenance is one of the important tasks which directly affects the firm's profit and production. For instance, a partial increment in retention rate has resulted in the maximized profits to a greater extent. The modifying costs are related to users switching behavior, and it is incurred by customers, firms, and both. Such costs are limited to monetary factors and it related to mental aspects that influence brand loyalty. Moreover, various switching costs are available like transaction cost, compatibility with a product, knowing to apply novel services, uncertainty regarding the supremacy of new services, and searching costs for novel producers. Hence, the costs are provided for massive corporations like the telecommunication sector [3]. Due to the existence of switching costs, firms ought to expend huge money and attract customers to retain users. Thus, maintaining the old users is better than earning new customers that is expensive and complicated. It is pointed out that, retention can be influenced by massive factors in the telecommunication sector, and some of them are network superiority, cost of a plan, photograph of a competitor, and mobile number constancy [4], and this literature concentrates in identifying the prominent aspects which affect the quality of CRM by applying machine learning (ML) methodologies.

Telecommunication firms have applied an extensive range of tricks like developing a collection of services, providing various ideas to satisfy the user demands, and offering discounts to retain user trustworthiness and inspire novel clients, wherein the profit can be enhanced gradually. Ironically, these tactics result in enhanced competition among several firms, and the prediction of user behavior is a complex operation. It prefers to offer a modified service to maintain the users preventing customer churn and also retain the increased competition. The norm churn means, loss of users who often change the supplier within a limited duration [5]. Certainly, customer retention is performed to reduce the customer churn which is a major problem in telecommunication industries, following which, the user demands are enhanced where the competition among the firms is also progressed in the telecommunication sector, also organizations apply the extensive range of procedures along with tactics.

One among them is the CRM model, which is highly concentrated in building the net profit along with key stockholders as well as users. The presence of digital systems deployment, information technologies, state-of-the-art ML as well as statistical methods offer better opportunities for telecommunication firms to learn the user requirements which is highly essential to present effectual CRM [6]. The economic factor of a company is affected by the users in the telecommunication industry so that even a small adjustment in financial objective has to be watched and monitored. Here, customer churn is caused because of the change in user behavior. But detecting customer

thoughts is not possible even by applying tactics and systematic ideas. It is essential to perform a novel study in learning customer behavior to find new factors of customer churn, and applying CRM would be highly advantageous in minimizing the churn rate effectively. Maintaining previous users is highly profitable instead of inspiring new customers [7]. Inspiring new users are expensive and no wonder why telecommunication service providers were trying to maintain the old customers. Thus, developing a précised churn prediction method and finding the users likely to churn are highly difficult operations.

In recent times, the ML method was employed in the CRM application. For instance, [8] identified the desired users at a stage, which begins to display a downward movement and a tendency to churn by applying a data gathered by actively reaching out to recent users and hearing the feedback. Firstly, the developers have defined and predict the downward trend propensity of users under the application of Decision Tree (DT) framework. Besides, semi-supervised learning, as well as casual inference, has been employed for identifying silent sufferers and golden set. [9] applies a supervised ML model to find the user who can be sustained in a mobile app-based business in which users subscribe for e-learning applications. The central premise of this study is to discover the aspects which are applicable to retain loyal customers and perform the marketing procedures accordingly.

This paper presents a new improved synthetic minority over-sampling technique (SMOTE) with optimal weighted extreme machine learning (OWELM) called the ISMOTE-OWELM model for CCP. The presented model includes preprocessing, balancing the unbalanced dataset, and classification. Firstly, customer data perform data normalization and class labeling. Afterward, the ISMOTE technique is employed to handle the imbalanced dataset. At last, the OWELM model is applied to compute the class labels of the applied data. The multi-objective rain optimization algorithm (MOROA) is used for two purposes: determining the optimal sampling rate of SMOTE and parameter tuning of WELM. In this study, ROA is chosen due to the following reasons. The ROA is a simpler and is an efficient searching technique. The presented model is proficient in the searching and identification of optimal solution in the large searching area with a reasonable amount of time. A series of simulations are carried out to ensure the ISMOTE-OWELM model against the CCP Telecommunication dataset.

Related works

Numerous models have been employed for predicting churn in the telecom industries. Mostly, ML and Data Mining (DM) methodologies were applied by several researchers in recent years. Followed by, several related works

have concentrated on using a single framework called data mining (DM) for knowledge extraction and the additional approaches have focused on relating numerous principles for churn prediction. Brandusoiu et al. [10] implied the latest module of DM to detect the churn for old customers by applying a dataset with call details of huge customers with features, and a reliant churn variable with 2 metrics namely, yes and no. Additionally, some features have data regarding the value input and output and voicemail of all users. Researchers have used principal component analysis (PCA) for computing dimensionality reduction. Also, three ML approaches were utilized they are neural networks (NN), Support Vector Machine (SVM), and Bayes Networks. The above-mentioned ML modalities are applicable for predicting the churn. Then, the developers have employed the area under curve (AUC) to estimate the model's efficiency. Finally, optimal values are gained from 3 ML models. Hence, a dataset applied in this work is minimum and it does not have any missing values. He et al. [11] developed an approach for churn prediction according to the NN model and overcome the issues of CRM in the large-scale Chinese telecom industry with a huge number of users. As a result, remarkable prediction accuracy has been attained to a greater extent.

Idris et al. [12] implied a framework based on genetic programming with AdaBoost and developed a churn prediction approach in telecommunications. A newly presented scheme is sampled under 2 standard datasets. Initially, Orange Telecom and secondly, cell2cell datasets; among these, maximum accuracy is gained from the cell2cell dataset even better by the first one. Huang et al. [13] analyzed the crisis of customer churn in a big data application. The key objective of developers was to ensure that big data is a capable one and maximize the churn prediction ability based on 3Vs namely, Volume, Variety, and Velocity of data. Random Forest (RF) method is applied and estimated with the help of AUC. Makhtar et al. [14] projected an approach for churn prediction under the employment of rough set theory in telecom services. The Rough Set classification approach has surpassed the previous models like Linear Regression, DT, and Voted Perception NN. Numerous authors have examined the crisis of imbalanced datasets in which churned customer classes are tiny when compared with active customer class labels, as it is the main problem in the churn prediction process. Amin et al. [15] preformed a comparison among six various sampling methods for oversampling about telecom CCP. The outcomes have showcased that the rules-generation relied on genetic algorithm (GA) has performed well than oversampling approaches. Burez and den Poel [16] learned the issues of imbalanced datasets in CCP schemes. Hence, the simulation outcome has represented that under-sampling frameworks have surpassed all the previous methodologies significantly.

The proposed ISMOTE-OWELM model

Figure 1 depicts the working process of the proposed ISMOTE-OWELM model. As shown in the figure, the input customer churn data perform data preprocessing, balancing the dataset, and classification. The MOROA is used for two purposes: determining the optimal sampling rate of SMOTE and parameter tuning of WELM.

Data preprocessing

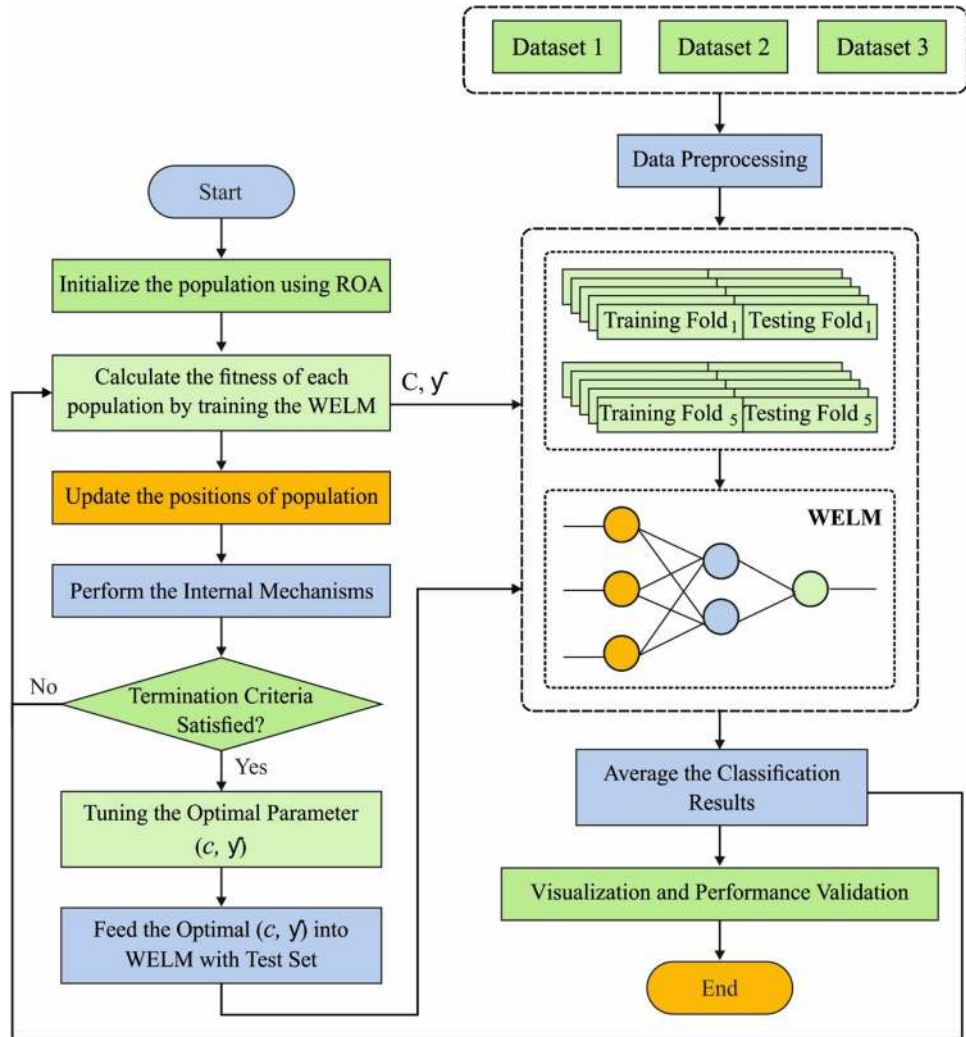
During the preparation step, it is discrete by size, the values that exist in all attributes of the dataset, after that allocated particular labels, e.g., Zero to Nine (0–9) feasible values, for all discrete group. In discretizing by size leads to choosing the numerical attributes to nominal attributes and grouped them into a particular size of bins. It divides the entire count of values in an attribute by the size of the bin. Finally, it created a particular list of values in several counts of groups of an attribute.

SMOTE for unbalanced dataset

The SMOTE is an over-sampling model [17], which intends to develop novel minority class samples by interpolating various original minority class samples. SMOTE randomly decides the k -nearest neighbors (kNN) of minority class instance, and performs a random interpolation of 2 instances for developing novel synthetic instance. In particular, the synthetic samples are produced as: apply the variations among an actual instance and the nearest neighbor, maximize these variations by a random number from 0 and 1, and include with the actual instance. It enforces the selection region of the minority class which is considered as a typical one. Hence, over-fitting issues are eliminated and decision boundary for minority class distributes further into the main class space. Based on SMOTE, numerous methods were presented for classifying the unknown data [18]. The previous SMOTE methods suffer from severe issues. It applies a similar sampling rate for instances of the minority class. Regardless, diverse samples comprise distinct roles in sampling and classification processes. The sampling rate is based on the sample instance, followed by the projected MOROA model to identify and apply the best sampling rates for different instances.

Choosing samples from minority class for over-sampling and fixing of sampling rates are relevant to the imbalanced degree of the dataset, entire distribution of samples, interior distribution of minority class samples, count of samples, count of sample attributes, and classes of attributes. It is a complex optimizing issue and resolved

Fig. 1 Block diagram of ISMOTE-OWELM model



by the computed numerical method. For reaching the maximum accuracy rate of minority class classification as well as optimal classification accuracy rate, diverse instances in the actual training set are correlated with sampling rates. The following ROA is applied for identifying actual sampling rates for different instances:

$$\begin{aligned} & \text{maximize} : y = f(x) \\ & \text{s.t.} : \min N \leq N_i \leq \max N; i = 1, 2, \dots, M; \\ & \text{where} : X = (N_1, N_2, \dots, N_M), \end{aligned} \quad (1)$$

where $f(X)$ implies the objective function, where the accuracy rate of minority class classification and entire classification of datasets; X denotes the decision vector, which means sampling rate; M refers the dimension of decision space, N_i means sampling rate of minority class sample x_i ; $\min N$ and $\max N$ are maximum and minimum bounds, correspondingly, of sampling rate N_i .

WELM-based classification

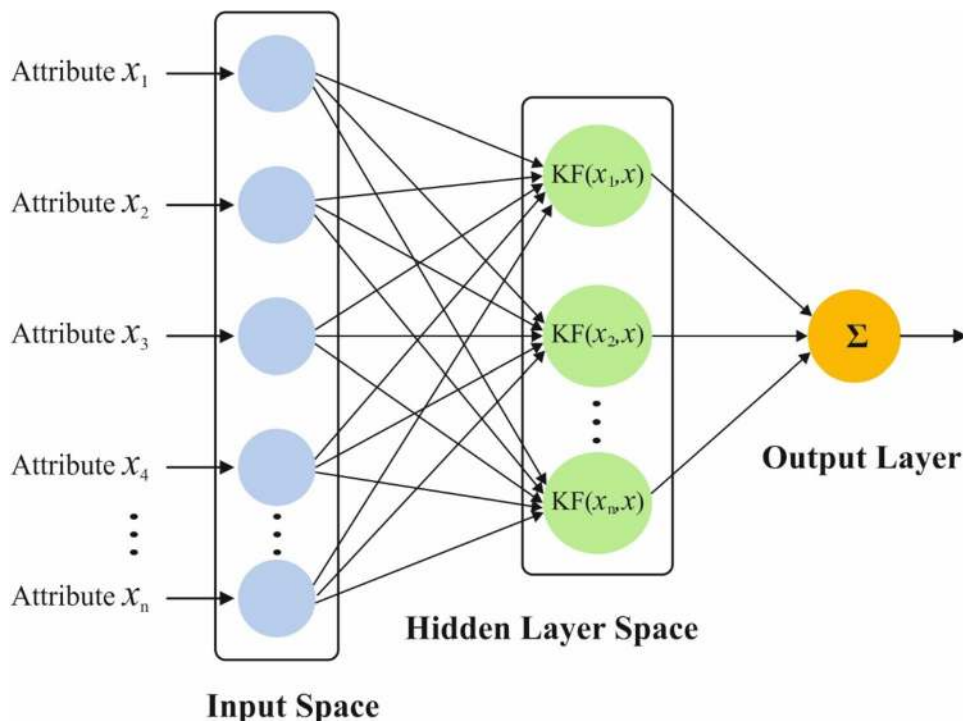
Extreme learning machine (ELM) is utilized to classify the balanced datasets, whereas WELM is utilized to classify the imbalanced datasets. The structure of ELM is shown in Fig. 2. So, this section describes the establishment of WELM [19]. The training dataset has N different instances $(x_i, z_i), i = 1, 2, \dots, N$.

The single hidden layer NN with L hidden layer nodes, which can be represented as:

$$\sum_{i=1}^L \beta_i \cdot l(w_i \cdot x_j + b_i) = z_j, j = 1, \dots, N, \quad (2)$$

where w_i implies a single hidden layer input weight, $l()$ defines the activation function, β_i denotes a resultant weight, and b_i represents the single hidden layer bias. Here, Eq. (2) is illustrated as:

Fig. 2 Structure of ELM



$$S\beta = T, \tag{3}$$

where S depicts the single hidden layer resultant matrix:

$$S(w_1, \dots, w_L, b_1, \dots, b_L, x_1, \dots, x_N) = \begin{pmatrix} l(w_L \cdot x_1 + b_1) & \dots & l(w_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ l(w_L \cdot x_N + b_1) & \dots & l(w_L \cdot x_N + b_L) \end{pmatrix}_{N \times L}. \tag{4}$$

Based on the Karush–Kuhn–Tucker theory, a Lagrangian factor is established for transforming the training of ELM into a dual problem. A resultant weight β is computed as follows:

$$\beta = S^T \left(\frac{1}{C} + SS^T \right)^{-1} T, \tag{5}$$

where C exhibits the regularization coefficient. So, the outcome function of the ELM classification model is illustrated as:

$$F(x) = s(x)S^T \left(\frac{1}{C} + SS^T \right)^{-1} T = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{1}{C} + \chi \right)^{-1} T, \tag{6}$$

where χ refers a kernel matrix which is computed as:

$$\chi = SS^T = s(x_i)s(x_j) = K(x_i, x_j). \tag{7}$$

It is seen from (5), that the hidden layer feature map $s(x)$ is independent of the result ELM classification, and the classification result is compared to kernel function $K(x, y)$. Afterward, the $K(x, y)$ applies the inner product, and so the hidden layer node count has no result on the output result. Thus, ELM does not require setting the input weights and offsets of the hidden layer. A kernel function of KELM is an RBF function, as defined below:

$$K(x, y) = \exp(-\gamma \|x - y\|^2). \tag{8}$$

Hence, KELM classification is carried out by 2 parameters: kernel function parameter γ and penalty parameter C . Based on the ELM advantages, the WELM with allocated weights to several samples can manage the imbalanced classification problem. Its resultant function is computed as:

$$F(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{1}{C} + W\chi \right)^{-1} WT, \tag{9}$$

$$W = \text{diag}(w_{ii}), i = 1, 2, \dots, N, \tag{10}$$

where W refers to a weight matrix. WELM has 2 weighting methods, i.e.,

$$w_{ii} = \frac{1}{\#(z_i)}, \tag{11}$$

$$w_{ii} = \begin{cases} \frac{0618}{\#(z_i)}, & \text{if } z_i > \bar{z}_i \\ \frac{1}{\#(z_i)}, & \text{otherwise} \end{cases}, \quad (12)$$

where $\#(z_i)$ is the count of samples going to class z_i , $i = 1, \dots, m$. m is the count of classes. \bar{z}_i is the average of the entire samples of every class. In WELM, the regularization coefficient C and bandwidth of RBF kernel γ are essential to determine the effectiveness of the performance of WELM.

ROA for sampling rate selection and parameter tuning

In this section, ROA is applied for the selection of optimal sampling rates of SMOTE and parameter tuning of the WELM model.

Inspiration

When rainfall starts, the droplets fall into the earth's surface. After some time, the droplets are joined with one another and it merges and moves on the surface based on the weight. Some other droplets move toward the consecutive droplet and merged with them and few might be exhausted or observed by the soil on the basis of soil features like the texture of soil surface, porosity, permeability, wettability, and so on. In addition, some soils are dissolved in the water, following which droplets dropped on flat region is absorbed by the soil completely and it has vanished from the inclined region and link other droplets for developing a stream. Certainly, few streams are connected with one another and change into a river. In case of any barrier in the path of streams or rivers, lakes are developed where the quantity of water represents the significance of droplets. Once the rain is stopped, streams and rivers are discharged to local lakes. Afterward, tiny lakes disappear because water evaporates in a lake or it is absorbed by the soil. Thus, the important lakes are sustained in groups according to the topology of the earth's surface and feature of a soil. Such lakes represent the local minima of ground surface as well as deeper lake implies the global minima.

When the rain type is changed, then there is a small change in the predefined procedure. For sample, in case of heavy rain with massive droplets, every droplet is linked with one another robustly with no absorption or exhaustion in a flood. At this point, the global minimum is predicted and the local minimum is correlated with one another because of a rainstorm. Followed by, if there is a light rain with the smaller droplet, every droplet is absorbed by the soil which leads to no stream development. Hence, it is analyzed that parameter tuning has vital significance at the time of applying ROA [20]. The movement

of a particle in the presented model is the same as gradient-related optimization approaches and traditional single-point models like hill-climbing (HC) and gradient-descent (GD) and rainfall optimization algorithm (RFO). These approaches modify a single parameter for all iterations for identifying whether the adjustment in this parameter enhances the cost function or not. Therefore, ROA applies a collection of replies where it moves toward the best one at the same time, following which these features change in all iterations.

Algorithmic steps of ROA

Here, the rain behavior is simulated as it is defined in the traditional section. Every solution to a problem could be referred to as a raindrop. Based on these issues, few points in the answer space are decided in a random fashion as raindrops fall on the earth surface. The major characteristic of a drop of rain is the radius. The radius of all raindrops could be limited as time passes and it is enhanced as a raindrop is linked with alternate drops. If the initial answer population is generated, the radius of every droplet is allocated in an arbitrary manner to a limited extent. Additionally, each droplet validates the neighborhood based on the size. Single droplets that are not yet linked just verifies the end limit of the place which has covered. In order to resolve the issues in dimensional space, each droplet is composed of n variable. Thus, in the initial step, the maximum and minimum limit of the variable is validated as the limits are computed by a radius of the droplet. Next, 2 endpoints of the variable are sampled and it is repeated till reaching the final variable. Afterward, the cost of the first droplet is upgraded by moving in a downward direction. It is performed for each droplet, and the cost, as well as position of every droplet, would be allocated. The radius of a droplet would be modified in 2 fashions:

When 2 droplets with radius r_1 and r_2 , then they are closer to one another with the typical area and it connects for developing larger droplet of radius R :

$$R = (r_1^n + r_2^n)^{1/n}, \quad (13)$$

where n implies the count of variables for all droplets. When a droplet with radius r_1 is not shifted, based on the soil features, that are showcased by α , water is observed by the soil.

$$R = (\alpha r_1^n)^{1/n}. \quad (14)$$

Obviously, α illustrates the volume of a droplet that has been absorbed in all iterations from 0 to 100 percent. Moreover, it describes the least values for droplets radius r_{\min} , where droplets with a minimum radius of that r_{\min} would be diminished.

As mentioned above, the population value might be reduced after some iterations and maximum droplets are

deployed with a larger domain of analysis. By enhancing the analysis process, the local searching potential of drops is maximized proportionally to the diameter of droplets. Thus, by maximizing the number of rounds, weaker droplets vanish or link with robust drops with the maximum domain of examination, and the initial population would be reduced intensively and find the correct answer(s). It is assumed that there are few variations among the newly presented optimization model in ROA and the currently implied search model deployed Rain Fall Algorithm (RFA) which is consolidated as:

- In ROA, the initial population number is modified after all iterations because of the connection of neighboring drops. It results in enhancing the searching capability of a model and reduces the optimization cost significantly.
- Once the size of a droplet is changed, the linking of the nearby droplets or adsorption by the soil is performed. This performance modifies the searching potential of every droplet and classifies the droplets.
- In RFA, and alternate searching models, every population is composed of neighbor points and the droplet is enhanced one step in a random fashion. Besides, every population identifies the optimal path to the lower point. Once the path is found, it is shifted in a downward direction iteratively by step and the cost function is reduced in a single iteration.

On the basis of the approximations and idealizations of the model, the rain method is depicted. In depth, tuning parameters of this method are initial raindrops number (population number), basic raindrops radius, and so forth, following which a value is allocated to all droplets on the basis of the cost function. Then, every droplet is moved in downward. Thus, near droplets combin with one another, which leads to enhanced results. If a droplet is terminated at the least point, the radius begins to decrease progressively causing the accuracy of the answer to be increased. Then, it is applicable to identify extrema points of the objective function.

Experimental validation

The presented model is simulated using Python 3.6 with keras and tensorflow packages. The performance of the presented ISMOTE-OWELM model is examined using three benchmark datasets namely 1–3. The first dataset includes 3333 samples with the presence of 21 features and 2 classes. Next, the second dataset comprises 7043 instances with the existence of 21 features with 2 class labels. Finally, the third dataset holds 100,000 samples with 100 features and 2 class labels. Besides, a detailed simulation analysis takes place

to ensure the goodness of the ISMOTE-OWELM model in terms of different performance measures. The related details of the dataset are provided in Table 1. For experimentation, tenfold cross-validation process is employed.

Table 2 and Figs. 3, 4 and 5 investigate the predictive performance of the ISMOTE-OWELM model in terms of different measures [24]. On the applied dataset-1, the proposed ISMOTE-OWELM model has attained better results in terms of different evaluation parameters. Under the execution run 1, the ISMOTE-OWELM model has resulted in a maximum sensitivity of 0.937, specificity of 0.929, accuracy of 0.931 and F-measure of 0.930. In addition, under the execution run 2, the ISMOTE-OWELM model has achieved a higher sensitivity of 0.948, specificity of 0.938, accuracy of 0.940 and F-measure of 0.931. Also, under the execution run 3, the ISMOTE-OWELM model has obtained better sensitivity of 0.939, specificity of 0.939, accuracy of 0.938 and F-measure of 0.932. Besides, under the execution run 4, the ISMOTE-OWELM model has depicted acceptable results with the maximum sensitivity of 0.939, specificity of 0.941, accuracy of 0.941 and F-measure of 0.941. Eventually, under the execution run 5, the ISMOTE-OWELM model has exhibited effective outcome with the sensitivity of 0.943, specificity of 0.948, accuracy of 0.946 and F-measure of 0.944.

On the applied dataset-2, the ISMOTE-OWELM model has portrayed optimal performance with respect to diverse performance measures. Under the execution run 1, the ISMOTE-OWELM model has led to a maximum sensitivity of 0.902, specificity of 0.926, accuracy of 0.919 and F-measure of 0.916. Similarly, under the execution run 2, the ISMOTE-OWELM model has attained a sensitivity of 0.913, specificity of 0.922, accuracy of 0.919 and F-measure of 0.918. Likewise, under the execution run 3, the ISMOTE-OWELM model has achieved a higher sensitivity of 0.927, specificity of 0.929, accuracy of 0.929 and F-measure of 0.922. Also, under the execution run 4, the ISMOTE-OWELM model has offered manageable outcome with the maximum sensitivity of 0.910, specificity of 0.919, accuracy of 0.914 and F-measure of 0.915. Finally, under the execution run 5, the ISMOTE-OWELM model has showed

Table 1 Dataset details

Description	Dataset 1	Dataset 2	Dataset 3
Number of samples	3333	7043	100,000
Number of features	21	21	100
Number of class	2	2	2
% of positive samples	14.49%	26.54%	49.56%
% of negative samples	85.51%	73.46%	50.43%
Data sources	[21]	[22]	[23]

Table 2 Performance evaluation of different runs on proposed ISMOTE-OWELM method

No. of runs	Sensitivity (%)	Specificity (%)	Accuracy (%)	F-Score (%)
Dataset-1				
Run-1	0.937	0.929	0.931	0.930
Run-2	0.948	0.938	0.942	0.931
Run-3	0.939	0.939	0.938	0.932
Run-4	0.939	0.941	0.941	0.941
Run-5	0.943	0.948	0.946	0.944
Average	0.941	0.939	0.940	0.935
Dataset-2				
Run-1	0.902	0.926	0.919	0.916
Run-2	0.913	0.922	0.919	0.918
Run-3	0.927	0.929	0.929	0.922
Run-4	0.910	0.919	0.914	0.915
Run-5	0.907	0.922	0.921	0.921
Average	0.912	0.924	0.920	0.918
Dataset-3				
Run-1	0.895	0.903	0.902	0.901
Run-2	0.874	0.905	0.904	0.903
Run-3	0.890	0.917	0.916	0.915
Run-4	0.875	0.918	0.918	0.916
Run-5	0.893	0.909	0.907	0.905
Average	0.885	0.910	0.909	0.908

operative outcome with the sensitivity of 0.907, specificity of 0.924, accuracy of 0.920 and F-measure of 0.918.

On the applied dataset-3, the ISMOTE-OWELM model has provided superior performance with respect to distinct methods. Under the execution run 1, the ISMOTE-OWELM model has showcased effective outcome with the maximum sensitivity of 0.895, specificity of 0.903, accuracy of 0.902 and F-measure of 0.901. Moreover, under the execution run 2, the ISMOTE-OWELM model has demonstrated a higher sensitivity of 0.874, specificity of 0.905, accuracy of 0.904 and F-measure of 0.903. Furthermore, under the execution run 3, the ISMOTE-OWELM model has out showed better sensitivity of 0.890, specificity of 0.917, accuracy of 0.916 and F-measure of 0.915. In line with, under the execution run 4, the ISMOTE-OWELM model has showcased acceptable results with the maximum sensitivity of 0.875, specificity of 0.918, accuracy of 0.918, and F-measure of 0.916. At last, under the execution run 5, the ISMOTE-OWELM model has displayed proficient results with the sensitivity of 0.893, specificity of 0.909, accuracy of 0.907, and F-measure of 0.905.

An average results analysis of the ISMOTE-OWELM model takes place on the test three datasets, as shown in Fig. 6. On the applied test dataset 1, the ISMOTE-OWELM model has attained a maximum average sensitivity of 0.941, specificity of 0.939, accuracy of 0.940, and F-measure of 0.935. Similarly, on the employed test dataset 2, the ISMOTE-OWELM model has obtained a higher average sensitivity of 0.912, specificity of 0.924, accuracy of 0.920

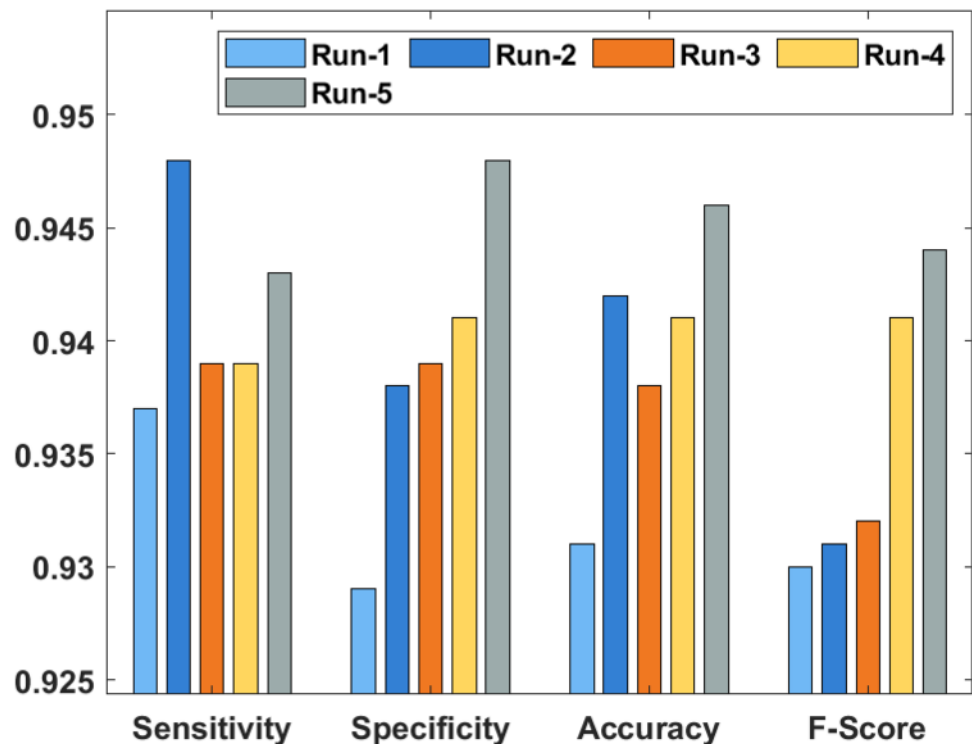
Fig. 3 Predictive performance analysis of ISMOTE-OWELM model on dataset-I

Fig. 4 Predictive performance analysis of ISMOTE-OWELM model on dataset-II

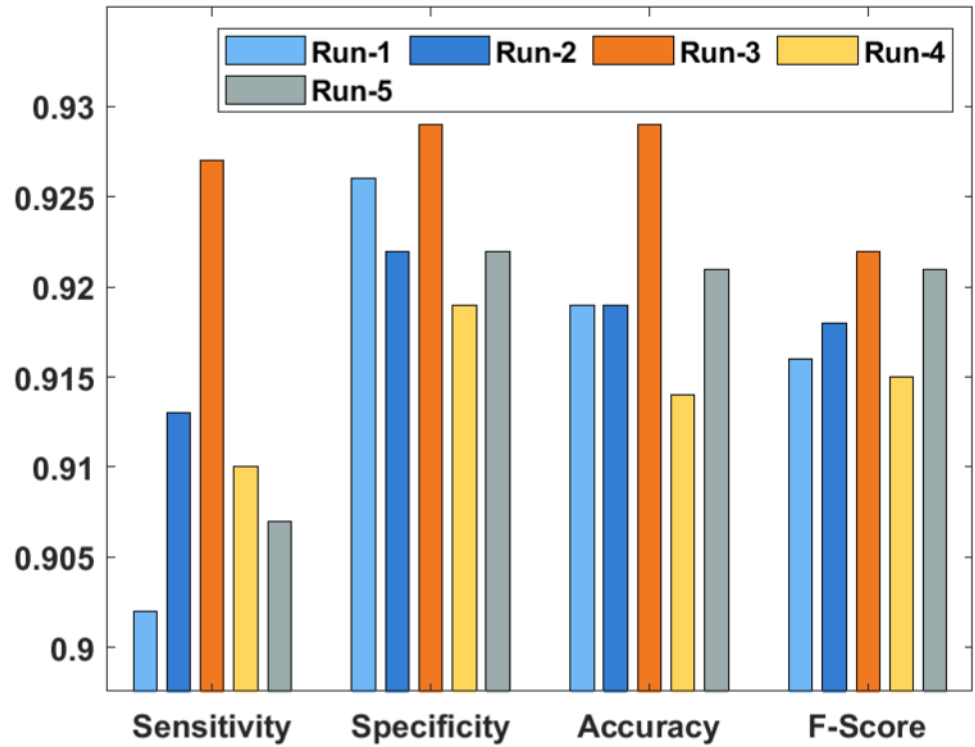
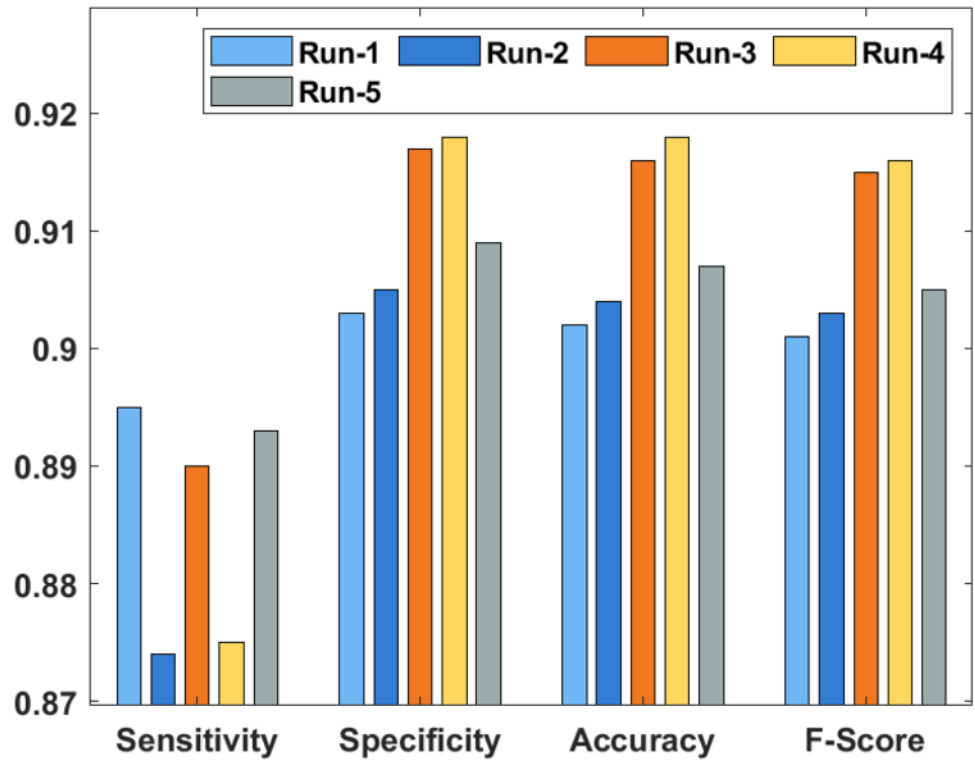


Fig. 5 Predictive performance analysis of ISMOTE-OWELM model on dataset-III



and F-measure of 0.918. Followed by, on the test dataset 3, the ISMOTE-OWELM model has resulted in an average sensitivity of 0.885, specificity of 0.910, accuracy of 0.909 and F-measure of 0.908.

Figures 7, 8 and 9 provide a detailed comparative predictive analysis of the ISMOTE-OWELM model on the applied 3 datasets. Figure 7 analyzes the results offered by the prediction performance of the ISMOTE-OWELM model on the

applied dataset 1. The results portrayed that the SVM model has failed to show better performance with the accuracy of 0.789 and F-measure of 0.763. At the same time, the PCPM model [25] has tried to exhibit certainly better results with the accuracy of 0.837 and F-measure of 0.838. Likewise, the LDT/UDT-1 model has attained an accuracy of 0.84 and F-measure of 0.579. Simultaneously, the LDT/UDT-2 model has resulted in a closer results with the accuracy of 0.84 and F-measure of 0.543. Along with that, the LDT/UDT-10 model has achieved even better results with the accuracy of 0.843 and F-measure of 0.56. Moreover, the LDT/UDT-8 model has attained effective performance with the accuracy of 0.846 and F-measure of 0.58. Furthermore, the LDT/UDT-10 model has achieved even better results with the accuracy of 0.843 and F-measure of 0.56.

On the other hand, the LDT/UDT-8, LDT/UDT-6, LDT/UDT-4, LDT/UDT-9, LDT/UDT-5, LDT/UDT-5, and LDT/UDT-7 models have demonstrated moderate and closer results in terms of accuracy and F-measure. Concurrently, it is exhibited that the WELM model has shown manageable results with the accuracy of 0.885 and F-measure of 0.882. In line with, the OWELM model has tried to surpass the earlier methods with the accuracy of 0.906 and F-measure of 0.904. In the same way, the SMOTE-OWELM model has showcased competitive results with the accuracy of 0.922 and F-measure of 0.93. But the proposed ISMOTE-OWELM model has depicted better results with the accuracy of 0.94 and F-measure of 0.935.

Figure 8 investigates the comparative analysis of the results provided by the outcome of the ISMOTE-OWELM model on the applied dataset 2. The simulation results depicted that the SVM model has shown worse results with the accuracy of 0.725 and F-measure of 0.731. Then, the LDT/UDT-1 model portrayed slightly higher results with the accuracy of 0.74 and F-measure of 0.618. Similarly, the LDT/UDT-8 model has reached an accuracy of 0.741 and F-measure of 0.613. Concurrently, the LDT/UDT-5 model has occasioned to nearer results with the accuracy of 0.742 and F-measure of 0.63. In line with, the LDT/UDT-7 model has attained even improved performance with the accuracy of 0.744 and F-measure of 0.62. Additionally, the LDT/UDT-9 model has reached operative performance with the accuracy of 0.747 and F-measure of 0.621. On the other hand, the LDT/UDT-6, LDT/UDT-10, LDT/UDT-4, LDT/UDT-3, and LDT/UDT-2 models have established reasonable and nearer results in terms of accuracy and F-measure. Also, the PCPM model has accomplished even improved results with the accuracy of 0.828 and F-measure of 0.931. Concurrently, it is displayed that the WELM model has revealed controllable results with the accuracy of 0.876 and F-measure of 0.872. Accordingly, the OWELM model has tried to exceed the previous models with the accuracy of 0.897 and F-measure of 0.894. Likewise, the SMOTE-OWELM model has performed competitive results with the accuracy of 0.918 and F-measure of 0.917. However, the ISMOTE-OWELM model has portrayed effective predictive outcome with the accuracy of 0.92 and F-measure of 0.918.

Fig. 6 Average results analysis of ISMOTE-OWELM model on different datasets

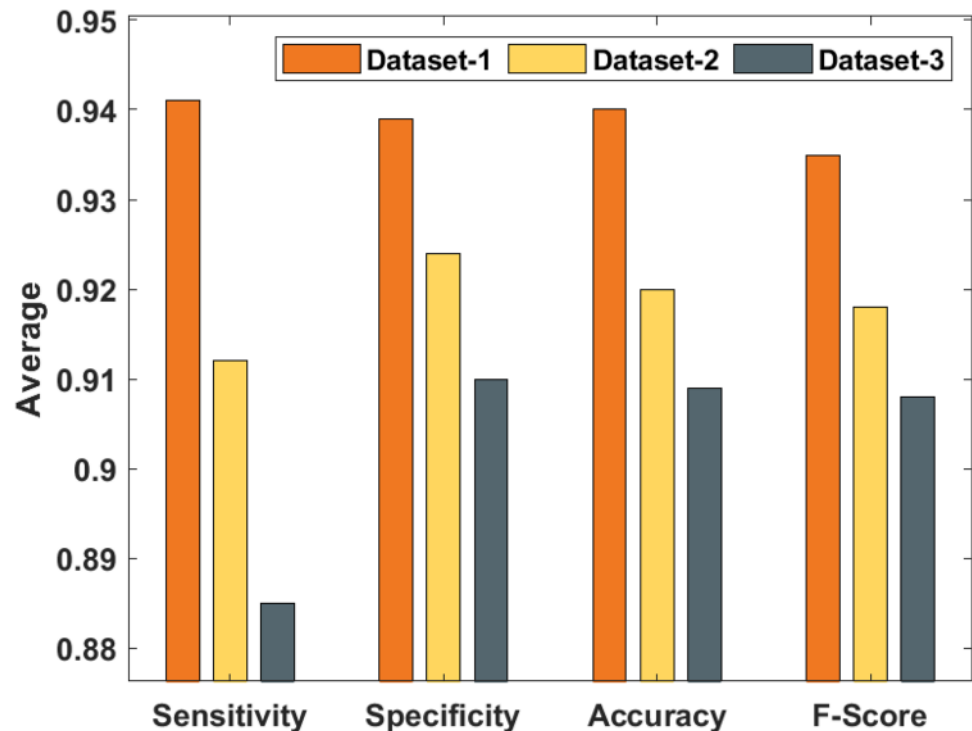


Fig. 7 Comparative results analysis of ISMOTE-OWELM model on dataset-I

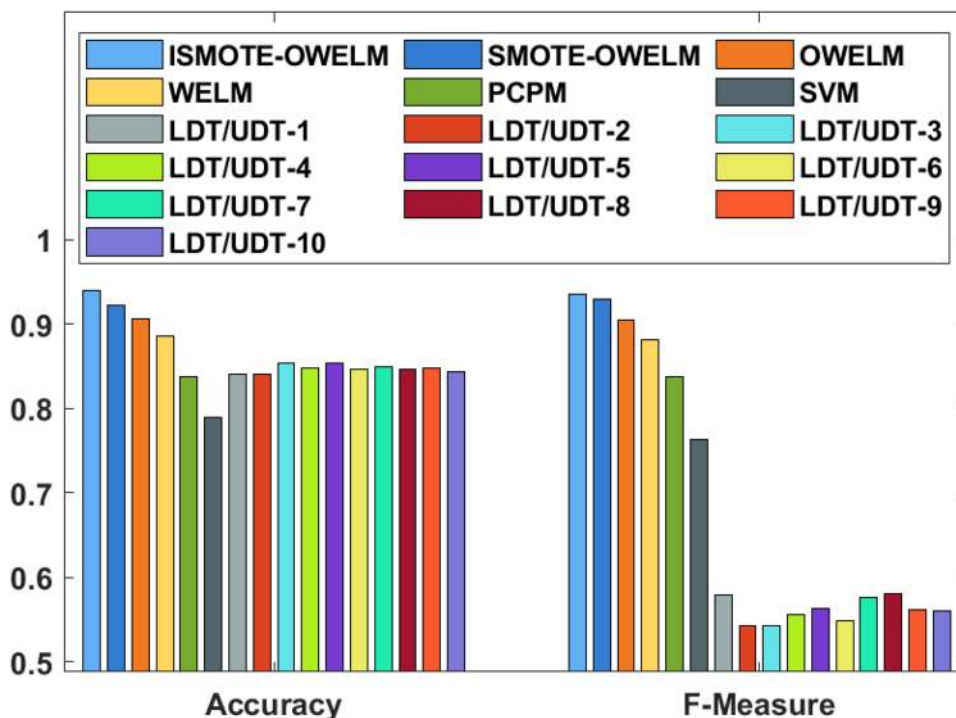


Fig. 8 Comparative results analysis of ISMOTE-OWELM model on dataset-II

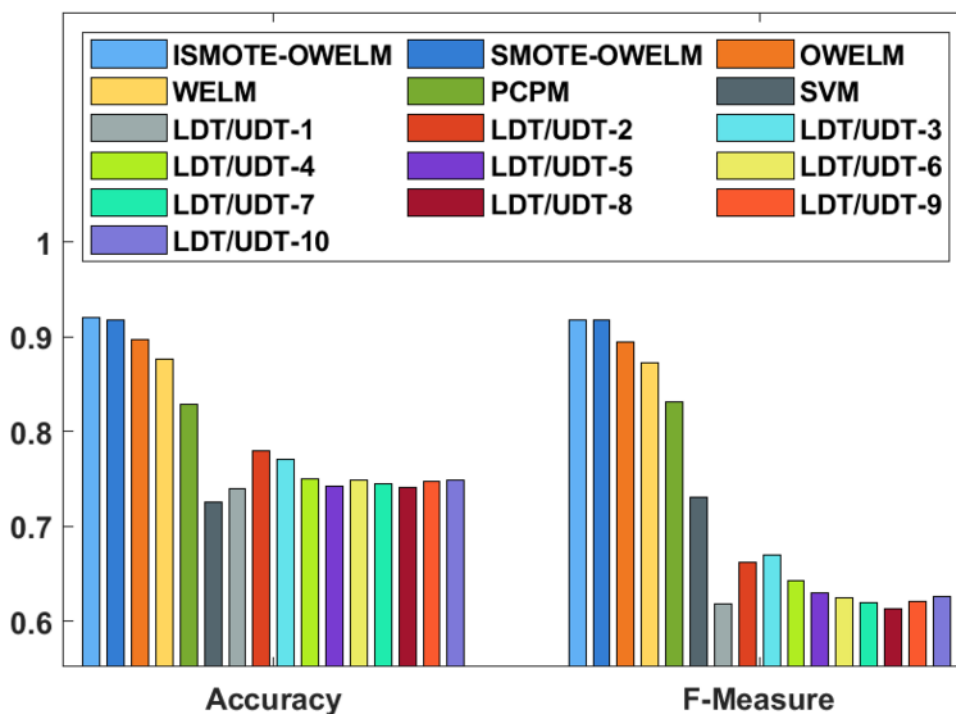
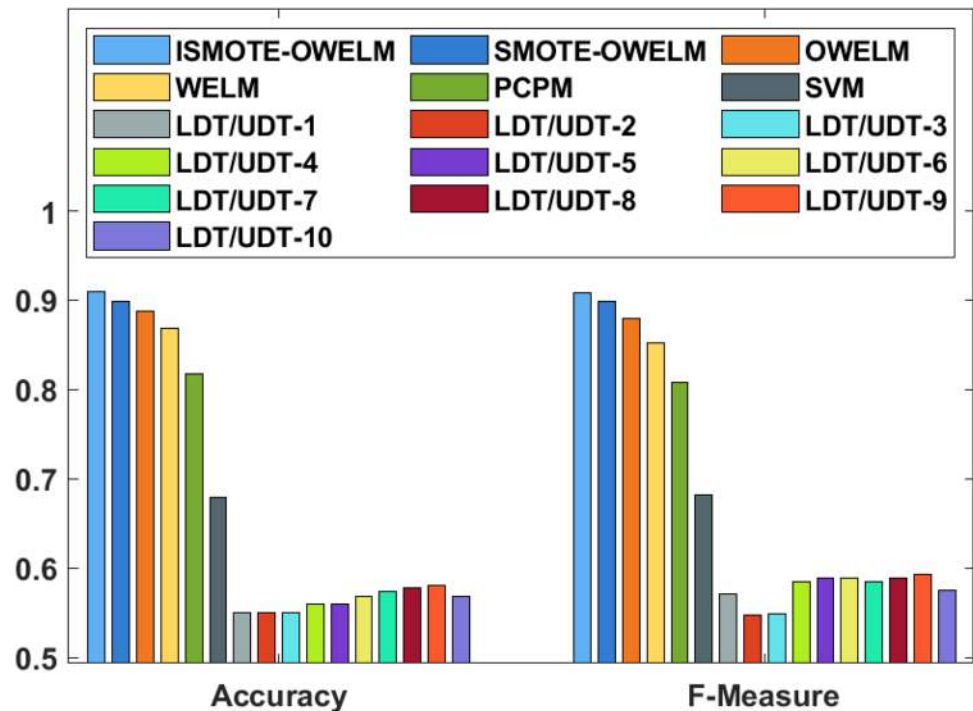


Figure 9 examines the comparative study of the ISMOTE-OWELM with the set of existing models on the applied dataset 2. The results depicted that the LDT/UDT-1 model is ineffective to attain better prediction with the accuracy of 0.55 and F-measure of 0.571. Next, the LDT/UDT-2 model has exposed somewhat effective outcome with the accuracy

of 0.55 and F-measure of 0.548. Similarly, the LDT/UDT-3 model has accomplished an accuracy of 0.55 and F-measure of 0.549. Instantaneously, the LDT/UDT-4 model has resulted to a closer results with the accuracy of 0.56 and F-measure of 0.584. In line with, the LDT/UDT-5 model has achieved even better results with the accuracy of 0.56 and F-measure

Fig. 9 Comparative results analysis of ISMOTE-OWELM model on dataset-III



of 0.588. Furthermore, the LDT/UDT-6 model has achieved proficient results with the accuracy of 0.568 and F-measure of 0.588. Additionally, the LDT/UDT-6, LDT/UDT-10, LDT/UDT-7, LDT/UDT-8, LDT/UDT-2, SVM and models have established moderate and closer results in terms of accuracy and F-measure. Also, the PCPM model has realized even better results with an accuracy of 0.818 and F-measure of 0.808. Concurrently, it is revealed that the WELM model has shown manageable results with the accuracy of 0.869 and F-measure of 0.852. In line with, the OWELM model has reported outstanding results over the earlier methods with the accuracy of 0.887 and F-measure of 0.879. Though the SMOTE-OWELM model has outperformed the previous compared methods, the ISMOTE-OWELM model has shown maximum predictive performance with the accuracy of 0.909 and F-measure of 0.908. The proposed model has achieved superior results due to the inclusion of ISMOTE to handle class imbalance problem and the efficient characteristics of ROA in the determination of optimal sampling rate. The presented model can be employed in real time e-commerce sites.

Conclusion

This paper has developed a novel ISMOTE-OWELM to determine the churners in the telecom sector. The presented model includes preprocessing, balancing the unbalanced dataset, and classification. Initially, the input customer churn data performed data preprocessing, balancing the dataset, and classification. The MOROA is used for two purposes: determining

the optimal sampling rate of SMOTE and parameter tuning of WELM. Initially, the customer data involve data normalization and class labeling. Then, ISMOTE technique is employed to handle the imbalanced dataset. At last, the WELM model is applied to determine the class labels of the applied data. The MOROA is used for two purposes: determining the optimal sampling rate of SMOTE and parameter tuning of WELM. A series of simulations were carried out and the ISMOTE-OWELM model has shown maximum predictive performance with the accuracy of 0.94, 0.92, 0.909 on the applied dataset I, II, and III, respectively. The experimental outcome stated that the performance of the ISMOTE-OWELM model over the compared methods. As a part of future extension, the performance can be further enhanced using different feature selection methodologies.

Acknowledgments The authors would like to thank the anonymous reviewers for their critical and constructive comments, their thoughtful suggestions have helped improve this paper substantially. Irina V. Pustokhina is thankful to the Department of Entrepreneurship and Logistics, Plekhanov Russian University of Economics, Moscow, Russia. Dr. K. Shankar would like to thank RUSA PHASE Dept. of Edn. Govt. of India.

Declarations

Conflict of interest The authors declare that they have no conflict of interest. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

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

References

- Nie G, Rowe W, Zhang L, Tian Y, Shi Y (2011) Credit card churn forecasting by logistic regression and decision tree. *Expert Syst Appl* 38(12):15273–15285
- Gustafsson A, Johnson MD, Roos I (2005) The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *J Mark* 69(4):210–218
- Grzybowski L (2008) Estimating switching costs in mobile telephony in the UK. *J Ind Compet Trade* 8(2):113–132
- Kim H-S, Yoon C-H (2004) Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommun Policy* 28(9–10):751–765
- Mock D (2011) Comparing carrier churn. <http://www.fool.com/investing/general/2007/06/15/wireless-smackdown-churn.aspx>. Accessed 09 Oct 2011
- Xu LD (2011) Enterprise systems: state-of-the-art and future trends. *IEEE Trans Industr Inf* 7(4):630–640
- Lu N, Lin H, Lu J, Zhang G (2014) A customer churn prediction model in telecom industry using boosting. *IEEE Trans Industr Inf* 10(2):1659–1665
- Hu K, Li Z, Liu Y, Cheng L, Yang Q, Li Y (2018) A framework in CRM customer lifecycle: Identify downward trend and potential issues detection. *arXiv preprint: arXiv:1802.08974*
- Jose J (2019) Predicting customer retention of an App-based business using supervised machine learning. Technological University Dublin
- Brandusoiu I, Todorean G, Ha B (2016) Methods for churn prediction in the prepaid mobile telecommunications industry. In: *International conference on communications*, p 97–100
- He Y, He Z, Zhang D (2009) A study on prediction of customer churn in fixed communication network based on data mining. In: *Sixth international conference on fuzzy systems and knowledge discovery*, vol. 1, 92–4
- Idris A, Khan A, Lee YS (2012) Genetic programming and adaboosting based churn prediction for telecom. In: *IEEE international conference on systems, man, and cybernetics (SMC)*, pp 1328–32
- Huang F, Zhu M, Yuan K, Deng EO (2015) Telco churn prediction with big data. In: *ACM SIGMOD international conference on management of data*, pp 607–18
- Makhtar M, Nafs S, Mohamed M, Awang M, Rahman M, Deris M (2017) Churn classification model for local telecommunication company based on rough set theory. *J Fundam Appl Sci* 9(6):854–868
- Amin A, Anwar S, Adnan A, Nawaz M, Howard N, Qadir J, Hawalah A, Hussain A (2016) Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. *IEEE Access* 4:7940–7957
- Burez D, den Poel V (2009) Handling class imbalance in customer churn prediction. *Expert Syst Appl* 36(3):4626–4636
- Chawla N, Bowyer K, Hall L, Kegelmeyer W (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357
- Jiang K, Lu J, Xia K (2016) A novel algorithm for imbalance data classification based on genetic algorithm improved SMOTE. *Arab J Sci Eng* 41(8):3255–3266
- Zhu H, Liu G, Zhou M, Xie Y, Abusorrah A, Kang Q (2020) Optimizing Weighted Extreme Learning Machines for Imbalanced Classification and Application to Credit Card Fraud Detection. *Neurocomputing*
- Moazzeni AR, Khomehchi E (2020) Rain optimization algorithm (ROA): A new metaheuristic method for drilling optimization solutions. *J Pet Sci Eng*, p 107512
- <http://www.sgi.com/tech/mlc/db/>. Accessed 30 Nov 2017 02:00 PM
- https://community.watsonanalytics.com/wp-content/uploads/2015/03/WA_FnUseC-Telco-Customer-Churn.xlsx. Accessed 30 Nov 2017
- <https://www.kaggle.com/abhinav89/telecom-customer/data> (Last Accessed 13 Dec, 2)
- Uthayakumar J, Vengattaraman T, Dhavachelvan P (2017) Swarm intelligence based classification rule induction (CRI) framework for qualitative and quantitative approach: an application of bankruptcy prediction and credit risk analysis. *J King Saud Univ-Comput Inf Sci*
- Shaaban E, Helmy Y, Khedr A, Nasr M (2012) A proposed churn prediction model. *Int J Eng Res Appl* 2(4):693–697

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