Intelligent Data Analytics using Hybrid Gradient Optimization Algorithm with Machine Learning Model for Customer Churn Prediction

Elvir Akhmetshin[1], Nurulla Fayzullaev[2], Elena Klochko[3], Denis Shakhov[4], Valentina Lobanova[5]

[1]Candidate of Economic Sciences, Associate Professor of Department of Economics and Management, Kazan Federal University, Elabuga Institute of KFU, Elabuga, Russia.
[2]Candidate of Economic Sciences, Associate Professor, Head of Department of Accounting and Audit, Urgench State University, Urgench, Uzbekistan.
[3]Doctor of Economic Sciences, Professor of Department of Management, Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia.
[4]Candidate of Sociological Sciences, Associate Professor of Department of Economics and Management, Khorezm University, Urgench, Uzbekistan.
[5]Candidate of Economic Sciences, Associate Professor of Department of Enterprise Economics, Regional and Personnel Management, Kuban State University, Krasnodar, Russia.
Emails: elvir@mail.ru; fayzullaev.n.b@mail.ru; klochko.e.n@yandex.ru; d.a.shkhv@gmail.com; valentina.v.lobanova@list.ru

Abstract

Intelligent data analytics for customer churn prediction (CCP) harnesses predictive modelling algorithms, machine learning (ML) techniques, and advanced big data analytics and also uncovers the underlying drivers and patterns of churn and detects customers at risk of churning. This business strategy help organization to implement retention efforts to decrease customer attrition and proactively detect at-risk customers. CCP allows businesses to take proactive measures such as targeted marketing campaigns, personalized offers, or enhanced customer service, to maintain valuable customer and decrease revenue loss. It is widely used in industries like telecommunications, subscription services, e-commerce, and finance to optimize customer retention strategies and enhance long-term profitability. ML algorithm can detect indicator and underlying trends that precedes churn by analyzing historical customer data, including transactional patterns, behaviors, demographics, and customer interaction. The study introduces Intelligent Data Analytics using Hybrid Gradient Optimization Algorithm with Machine Learning (IDA-HGOAML) Model for Customer Churn Prediction. The main intention of IDA-HGOAML method focuses on the prediction and classification of customer churns and non-churns. To do so, the IDA-HGOAML technique initially undergoes data pre-processing using Z-score normalization. The IDA-HGOAML model makes use of equilibrium optimization algorithm (EOA) for the feature selection (FS). Besides, the churn prediction method is implemented by the convolutional autoencoder (CAE) model. Finally, the HGOA is exploited for the optimal hyperparameter selection of CAE model, thereby enhancing the prediction results. A widespread experimental analysis were performed to validate the enhanced efficiency of the IDA-HGOAML method. The extensive outcomes indicated the improved prediction results of the IDA-HGOAML method over existing techniques in terms of different measures.

Keywords: Intelligent systems; Data analytics; Customer churn prediction; Feature selection; Machine learning

1. Introduction

The advancement and globalization of the telecommunications sector, drastically increase the competition and escalate the several operators in the market [1]. In the modern era, it is essential to periodically increase the profit, for that different strategies such as up-selling existing customers, acquiring new customers and extending the

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

retention period of customer has been introduced [2]. Among them, retention of existing customers is more cost-effective than others. Companies should reduce the potential customer churn to adopt the third strategy, viz., customer movement from one service provider to another [3]. The most important reason for churn is the dissatisfaction with support system and consumer services. The solution to the given problem is to predict the customers who are easily affected by churning [4]. Customer churn prediction (CCP) using big data is an area of research within the machine learning (ML) technique that works to categorize different kinds of customers into churning or non-churning customers [5]. Large number of studies have created prediction models based on the statistical and data mining approaches (ML model) namely neural network (NN), decision trees (DTs), linear regression, logistic regression (LR), random forest (RF), deep neural network (DNN), and support vector machines (SVM). The major objective of CCP is to establish strategies for customer retention [6]. Also, the risk of customer churn increases exponentially together with increasing competition in the market for providing services. Thus, establishing strategies to monitor non-churners (loyal customers) has become essential [7]. The customer churn model aims to forecast the customers who voluntarily leave and recognize premature churn signals.

Over the past decades, the increase in the availability of customer data coupled with the low cost of computational infrastructure and data storage fostered the usage of supervised ML techniques for predicting customer churn [8]. The ability to capture complex, non-linear dependencies between the churn events and features and to scale well with high dimensional data (for example, customer features and an increasing number of customers) makes the ML model the tool of choice for CCP in applications of academic and industry [9]. The latest advances in DNN architecture for sequential data can potentially address the challenges intrinsic to shallow ML models for exploiting time-varying data. Given a comprehensive space of architectural choices, the architecture of the DNN-based CCP technique that adapts both time-invariant and time-varying features is not direct [10]. Furthermore, based on gradient boosting models, contemporary empirical studies in classification-based tasks in the financial sector have suggested that the DL model may not outperform simple alternatives.

This study introduces Intelligent Data Analytics using Hybrid Gradient Optimization Algorithm with Machine Learning (IDA-HGOAML) Model for Customer Churn Prediction. The main intention of IDA-HGOAML method focuses on the prediction and classification of customer churns and non-churns. To do so, the IDA-HGOAML technique initially undergoes data pre-processing using Z-score normalization. The IDA-HGOAML model makes use of equilibrium optimization algorithm (EOA) for the FS process. Besides, the churn prediction method is implemented by the convolutional autoencoder (CAE) model. Lastly, the HGOA is employed for the optimum hyperparameter selection process of the CAE model, thereby improving the prediction results. A detailed set of experiments were carried out to validate the higher efficiency of the IDA-HGOAML method.

2. Related Work

Jajam et al. [11] advised a recent RNN and Stacked Bi-directional Long & Short Term Memory (SBLSTM) method for the Arithmetic Optimization Algorithm (AOA) in CCP. At an initial stage, the AOA method executes pre-processing. Additionally, the SBLSTM-RNN technique is utilized in order to differentiate between non-churning and churning customers. An optimum Hyperparameters tuning procedure employing an Improved Gravitational Search Optimization Algorithm (IGSA) is applied to improve the CCP model results. Amatare and Ojo [12] proposed a CNN technique to predict customer churning performance, therefore two datasets have been created. For the development and implementation, the Python programming language through the anaconda distribution was utilized. The IDE learning choice is the Jupyter Notebook. To attain a like-for-like comparison, two models were established namely CNN and Multi-layer Perceptron (MLP) models. Gabhane and Aslam Suriya [13] employed DL methodologies like Artificial Neural Networks (ANN) and CNN in the proposed development method. The information is gathered from the website of Kaggle dataset. Then, the dataset is pre-transformed by employing numerous methods and the essential features are removed. After the feature extraction, DL methods like ANN and CNN for developing the model.

In [14], mainly focuses on the Stacked DL development with Wind Driven Optimization based Business Intelligence for CCP technique. The projected method is considered an intelligent structure which implements the Golden Sine Algorithm (GSA) established FS technique to develop salient features. Furthermore, the Stacked GRU (SGRU) methods are employed to predict customer churns. In [15], a CCP technique is proposed by using the ML approach in the CC environment. This one employs three kinds of stages likely preprocessing, data collection, and Adaptive Gain with Backpropagation Neural Networks (AGBPNN), which is referred to as P-AGBPNN. At the initial stage, by using numerous IoT devices, customer data is collected. The IoT devices send the collected information to the Cloud Data Server (CDS). Next, the pre-processing step takes place. After that, the developed technique is performed on the cloud itself. Pekel Ozmen and Ozcan [16] initially performed a CNN method on an arithmetic data set for a customer churn classification in transactions. Afterwards, due to several data losses in data transformations, a novel Extended Convolutional DT technique (ECDT) was mainly designed

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

to improve the CNN model. At last, an ECDT-GRID were designed to develop the classification accuracy of ECDT by performing grid search optimization.

In [17], the research projected a churn classification which is based on the enhanced DL classification algorithm which is built into the spark design. Then the ODL model is recognized via the optimum learning of the DCNN by employing developed Firefly-Spider Optimization (FSO), which is combination of the firefly optimization (FA) algorithm as well as Spider Monkey Optimization (SMO). Adnan et al. [18] develop an adaptive learning method for the disconcerting issue of CCP by utilizing the classification model with a GA-based feature weighting method. Moreover, the suggested technique performance is estimated on publicly accessible datasets namely IBM Telco, Cell2Cell and BigML Telco churn which highly improves the prediction performance when it is compared to the base classification model.

3. The Proposed Model

In this study, an IDA-HGOAML method was introduced for the prediction of customer churns. The primary goal of IDA-HGOAML model focuses on the prediction and classification of customer churns and non-churns. To do so, the IDA-HGOAML technique incorporates different processes namely data pre-processing using Z-score normalization, EOA-based FS, CAE classification, and HGOA-based parameter tuning. The working flow of the IDA-HGOAML methodology is depicted in Fig.1.

A. Pre-processing

The Z-score normalization is primarily applied to normalize the data inputted into scalar format [19]. Here, Z-score normalization reduce the variance in large datasets and accelerate the training process. It can be defined as follows.

$$z = \frac{(x - \mu)}{\sigma} \tag{1}$$

Where z indicates the z-score normalization, x shows the value, the standard deviation of sample is represented by the symbol σ and the sample mean is μ.

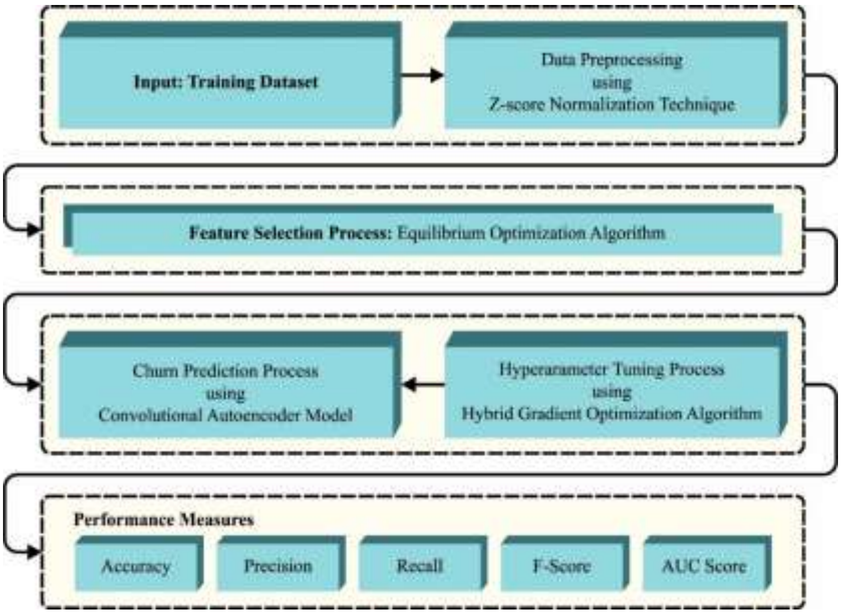

Figure 1: Workflow of IDA-HGOAML algorithm

B. Feature Selection Module

The EOA is used for electing an optimum set of features. EOA is a new promising algorithm based on the control volume mass balance model [20]. Like other techniques, EOA provides more accurate results and is used to attain the optimum hub height of the wind turbines. It estimates the equilibrium and dynamic states, where, all the particles with their concentration act as a searching agent. As shown in Eq. (2), the concentration is updated randomly for the search agent concerning the equilibrium state (optimum result), and then the equilibrium candidates (best solution) can be obtained.

$$CON = COeq + (CON1 - COeq) \times f + \frac{g}{\lambda \times \mathrm{vo}}(1 - f) \tag{2}$$

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

In Eq. (2), the initial concentration is $CON1$, vo represents the control volume, the concentration in the equilibrium state can be defined as $COeq$, f refers to exponential term, λ indicates the turnover rate, and g denotes the mass generation rate.

The three important stages of the EOA technique can be discussed in the following:

Initially, the calculation of function that initializes the population. It relies on the dimension and the particle number with random initialization in the examined space as shown in Eq. (3). The minimal and maximal values of the dimension are $Cmin$ and $Cmax$. The concentration vector of j^{th} particles is referred to as Cin_j, K represents the particle count, and $rand_j$ is the random vector.

$$Cin_j = Cmin + rand_j(cmax — Cmin)\ forj = 1,2,\cdots K \quad (3)$$

The next stage deliberates the equilibrium pooling and the candidate nominated from the population. Besides the particles which have an average of the best element, the topmost four particles at balance candidate are used. By using these five equilibrium candidates, the exploration and exploitation phases can be optimized by the following equation:

$$Cpool = [COeq1, COeq2, COeq3, COeq4, COeq(ave)] \quad (4)$$

In the last stage, the EOA is used to update the concentration with moderate equilibrium between exploration and exploitation. The equation for updating the EOA rule is given in the following.

$$\overrightarrow{CON} = \overrightarrow{COeq} + \left(\overrightarrow{CON} - \overrightarrow{COeq}\right) \times \vec{f} + \frac{\vec{g}}{\vec{\lambda} \times vo}\left(1 - \vec{f}\right) \quad (5)$$

In Eq. (5), $\overrightarrow{CON}$ denotes the updated position.

The objective is combined into single objective formula thereby a preset weight detects the importance of the objectives [21]. In this technique, we adopt an FF that incorporates both objectives of FS using the following expression:

$$Fitness(X) = \alpha \cdot E(X) + \beta * \left(1 - \frac{|R|}{|N|}\right) \quad (6)$$

In Eq. (6), the fitness value of subset X is Fitness(X), the classifier error rate by using the selected feature in X subset is $E(X)$, the amount of features selected and the amount of original features in the dataset are $|R|$ and $|N|$, the weights of classifier error rate and the reduction ratio are α and β correspondingly, where $\alpha \in [0,1]$ and $\beta = (1 - \alpha)$.

C. Churn Prediction using CAE Model

To predict the customer churns, the CAE model is applied. An autoencoder (AE) is a feedforward NN (FFNN) that will recreate the input at output in specific conditions [22]. That is to say, the AE attempts to estimate $h_w(x) \approx x$ for the input vector of x. An unsupervised pretraining model was introduced to initialize the weight and bias of the AE which is very efficient when few labelled training sample was available. First, AE implements unsupervised pre-training via encoding and later decodes the input. Fig. 2 demonstrates the framework of CAE.

The encoder calculates a nonlinear mapping of the input for the input vector x, as follows:

$$e_i = \sigma(Wx_i + b). \quad (7)$$

Now, the nonlinear activation function is represented by σ, W and b are the weight and bias of the encoder.

$$z_i = \sigma\left(\widetilde{W} e_i + \tilde{b}\right). \quad (8)$$

In Eq. (8), the weight and bias vectors of the decoder is represented as $\widetilde{W}$ and $\tilde{b}$. In the unsupervised pre-training, the network attempts to reduce the reconstructed errors by modifying the weight and bias $\theta = \left[W, b, \widetilde{W}, \tilde{b}\right]$.

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}(x_i - z_i)^2 \quad (9)$$

A sparsity parameter is added to the cost function for preventing the network from learning the identity function. The advantages of the convolution filter in CNN are combined with the unsupervised pretraining of AE. The

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

encoder has convolution layer and the decoder has deconvolution layer rather than FC layer. Deconvolution filters might be learned from scratch; or, they might be transposed versions of convolution filters. Furthermore, the deconvolution layer should be followed by the unpooling layer. The unpooling process can be done by storing the location of the maximal values, which preserves the value of this location.

Spatial locality can be retained by integrating the convolution function at all the neurons. Therefore, for the input matrix P, the encoder calculates

$$e_i = \sigma(P * F^n + b) \tag{10}$$

In Eq. (10), σ is an activation function, $*$ is a $2D$ convolution, F^n refers to $the\ n^{th} 2D$ convolutional filter and b is an encoder bias. Zero padding is used to the input matrix P to preserve spatial resolution. Next, the reconstruction is attained by the following equation:

$$z_i = \sigma\left(e_i * \tilde{F} + \tilde{b}\right). \tag{11}$$

Now, z_i refers to the reconstruction of i^{th} input, $\tilde{F}^n$ is n^{th} $2D$ convolution filters in the decoder and b is bias of the decoder.

$$E(\theta) = \sum_{i=1}^{m} (x_i - z_i)^2 \tag{12}$$

At the network end, the decoder part of the network is removed. The FC layer along with softmax classifier is added after unsupervised pretraining of the deconvolution and unpooling layers.

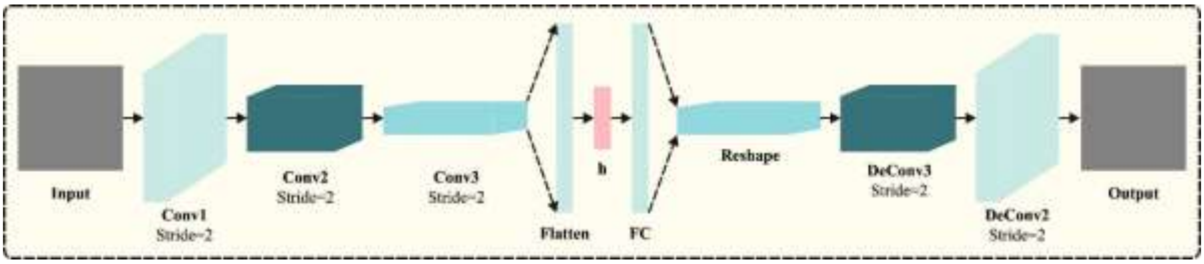

Figure 2: Framework of CAE

D. Hyperparameter Tuning Module

At last, the parameters related to the CAE model can be chosen by HGOA. The standard approaches of resolving the maximum likelihood contain the Levenberg-Marquardt, Gauss-Newton and Newton-Raphson approaches [23]. The Gauss-Newton gradient optimizer method and genetic algorithm are combined by using an HGOA.

Initially, the HGOA carries out GA without previous knowledge for primary iteration, and the next optimizer utilises the Gauss-Newton method. As a huge primary error can affect scattered parameter evaluations, the GA can be employed at a primary moment of computation to decrease the sensitivity to the primary values. Besides, the Gauss-Newton has faster convergence rate. But it creates incorrect gradient data and effect divergence once it can be away from the actual solution.

Because of the inherent arbitrariness of a GA and exploration for better performance, this method has maximum likelihood of determining the global optimum value of problem. The GA primary needs to generate an initial population. The final is arbitrarily chosen on the upper and lower limitations, and it can be expressed by the subsequent matrix:

$$Initial\ population\ = \begin{bmatrix} \theta_1^1 & \theta_2^1 & \dots & \theta_{p_i}^1 \\ \theta_1^2 & \theta_2^2 & \dots & \theta_{p_i}^2 \\ \vdots & \vdots & \dots & \vdots \\ \theta_1^{p_j} & \theta_2^{p_j} & \dots & \theta_{p_i}^{p_j} \end{bmatrix} \tag{13}$$

Whereas p_i denotes the count of model parameters, p_j signifies the count of the parameter vector, $\theta_{J_i}^{p_i}$ in all the rows defines the model parameter p_i from the parameter vector p_j, and all the rows of population matrix define the gene series of individuals.

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

Utilizing gene manipulation, the GA elects individuals with superior fitness. The population size can be massive sufficient for all the generations to take a maximum likelihood of seeking global optimal.

The fitness of all the individuals from the population can evaluated by the main function. The procedure of creating a novel population by executing 3 gene manipulations on all the individuals of existing population is provided as:

Select: the best individuals can elected in the existing population and copied to the next generation.

Crossover: the 2 individuals were arbitrarily elected in the existing population. The location of crossover point from the gene can arbitrarily be chosen for simulating the reproduction phenomenon from the development procedure, so, that individuals with novel genes can achieved.

Mutate: the existing gene arbitrarily chooses the mutation point for changing with a smaller probability, and next creates a novel individual.

These phases can repeated till the end state is met. While the GA utilizes likelihood search technique, the transfer mode and connection from single search point to the other are undefined which causes inconsistent performances in all the iterations. Therefore, the GA could not utilized only for estimating the PEC model parameters dependent upon testing data. Additionally, to achieve a correct parameter estimate, the performance of GA has been employed as a primary parameter evaluated using Gauss-Newton optimizer method. This gradient-based optimizer method is a typical approach for resolving MLE optimizer problem. The iterative formula is expressed as:

$$\hat{\theta}(i+1)=\hat{\theta}(i)+\Delta\hat{\theta}, \Delta\hat{\theta}=-M_{\theta=\theta(i)}^{-1}g_{\theta=\theta(i)} \tag{14}$$

Whereas θ parameter vector at step i is denoted as $\hat{\theta}(i)$, and $g_{\theta=\theta(i)}$ implies the gradient of probability function comparative to the model parameter that is achieved by computing the Jacobian matrix comparative to θ.

Furthermore, to avoid the Gauss-Newton optimizer method in terminating and diverging the iteration, this method can transferred to Nelder-Mead simplex approach and this approach is a multi-dimensional direct searching method for local optimizer:

$$\hat{\sigma}=\frac{1}{m}\sum_{k=1}^{m} v^2\,(k) \tag{15}$$

It is realized that the variance evaluation $\hat{\sigma}$ is connected to θ model parameter vector. For the actual optimizer, the relaxation system was frequently utilized for optimizing the main function. It contains an alternate estimate of the variance σ_n^2 and the parameter vector θ. The variance estimate $\hat{\sigma}$ is attained depending on a set model parameter vector θ, by setting $\sigma_n^2=\hat{\sigma}$. The variable that is optimized then develops θ. Based on the θ existing model parameter vector, variance evaluation $\hat{\sigma}$ is computed once the iteration step is smaller. At last, these stages can repeat still until the convergence condition is met.

The fitness selection is a most important factor in the HGOA technique. The encoding solution is utilized to estimate the goodness of solution candidate. The accuracy values are the most important condition used for designing the FF.

$$Fitness \;=\; \max{(P)} \tag{16}$$

$$P=\frac{TP}{TP+FP} \tag{17}$$

Here TP and FP are the true and false positive values.

4. Performance Validation

The CCP outcomes of the IDA-HGOAML method can be validated on the Telecom churn dataset [24], comprising 20 features and 3333 samples as represented in Table 1. The IDA-HGOAML method has selected 12 features from the applied dataset.

Table 1: Details on database

Classes	No. of Samples
Non-Churn	2850
Churn	483

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

Total Samples	3333

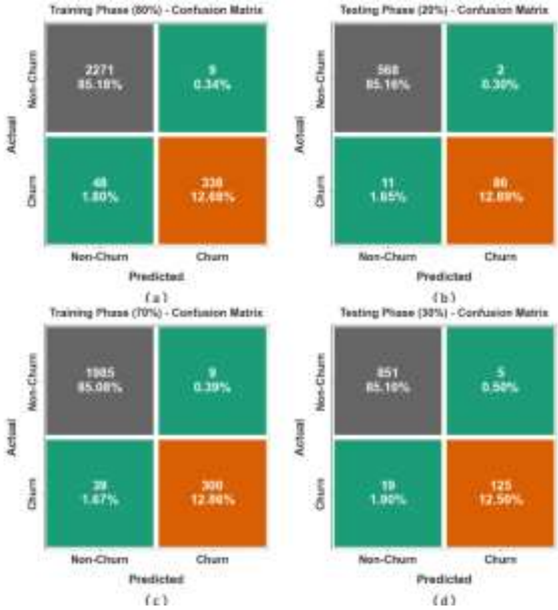

Figure 3: Confusion matrices of (a-c) TRSPH phase (TRPH) of 80% and 70% and (b-d) TS phase (TSPH) of 20% and 30%

The confusion matrices accomplished by the IDA-HGOAML method on 80:20 and 70:30 of the TRPH/TSPH is demonstrated in Fig. 3. The outcomes indicated the efficient detection and classification of 2 class labels.

The CCP results of the IDA-HGOAML technique on the 80:20 of TRPH/TSPH are given in Table 2 and Fig. 4. The results highlighted that the IDA-HGOAML method properly recognized the non-churn and churn samples. With 80% of TRPH, the IDA-HGOAML method provides average $accu_y$, $prec_n$, $reca_l$, F_{score}, and AUC_{score} values of 93.59%, 97.67%, 93.59%, 95.49%, and 93.59%, correspondingly. Meanwhile, with 20% of TSPH, the IDA-HGOAML technique provides average $accu_y$, $prec_n$, $reca_l$, F_{score}, and AUC_{score} values of 94.15%, 97.91%, 94.15%, 95.92%, and 94.15%, correspondingly.

Table 2: CCP outcome of IDA-HGOAML method on 80:20 of TRPH/TSPH

Classes	$\boldsymbol{Accu_y}$	$\boldsymbol{Prec_n}$	$\boldsymbol{Reca_l}$	$\boldsymbol{F_{Score}}$	$\boldsymbol{AUC_{Score}}$
TRPH (80%)					
Non-Churn	99.61	97.93	99.61	98.76	93.59
Churn	87.56	97.41	87.56	92.22	93.59
Average	93.59	97.67	93.59	95.49	93.59
TSPH (20%)					
Non-Churn	99.65	98.10	99.65	98.87	94.15
Churn	88.66	97.73	88.66	92.97	94.15
Average	94.15	97.91	94.15	95.92	94.15

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

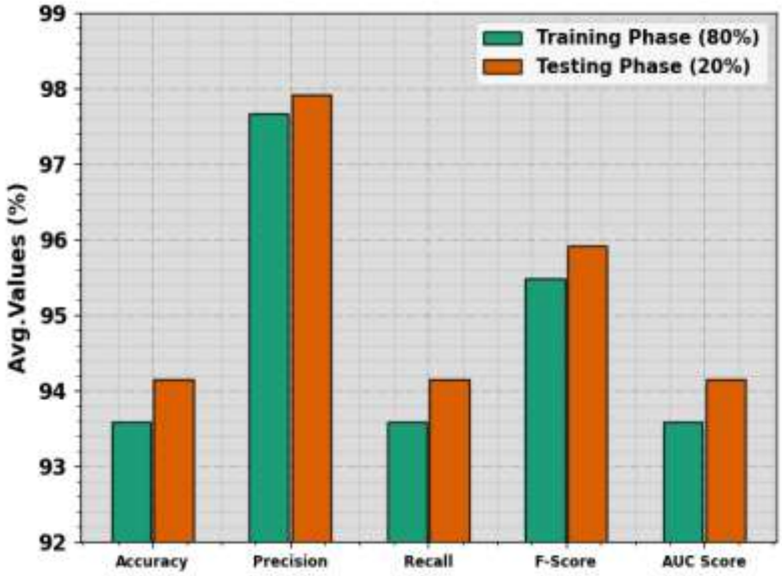

Figure 4: Average of IDA-HGOAML method on 80:20 of TRPH/TSPH

The CCP results of the IDA-HGOAML system on the 70:30 of TRPH/TSPH are given in Table 3 and Fig. 5. The outcomes highlighted that the IDA-HGOAML system properly recognized the non-churn and churn samples. With 70% of TRPH, the IDA-HGOAML system shows average $accu_y$, $prec_n$, $reca_l$, F_{score}, and AUC_{score} values of 94.02%, 97.58%, 94.02%, 95.70%, and 94.02%, correspondingly. Meanwhile, with 30% of TSPH, the IDA-HGOAML method provides average $accu_y$, $prec_n$, $reca_l$, F_{score}, and AUC_{score} values of 93.11%, 96.98%, 93.11%, 94.93%, and 93.11%, correspondingly.

Table 3: CCP outcome of IDA-HGOAML method on 70:30 of TRPH/TSPH

Classes	$\boldsymbol{Accu_y}$	$\boldsymbol{Prec_n}$	$\boldsymbol{Reca_l}$	$\boldsymbol{F_{Score}}$	$\boldsymbol{AUC_{Score}}$
TRPH (70%)					
Non-Churn	99.55	98.07	99.55	98.81	94.02
Churn	88.50	97.09	88.50	92.59	94.02
Average	94.02	97.58	94.02	95.70	94.02
TSPH (30%)					
Non-Churn	99.42	97.82	99.42	98.61	93.11
Churn	86.81	96.15	86.81	91.24	93.11
Average	93.11	96.98	93.11	94.93	93.11

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

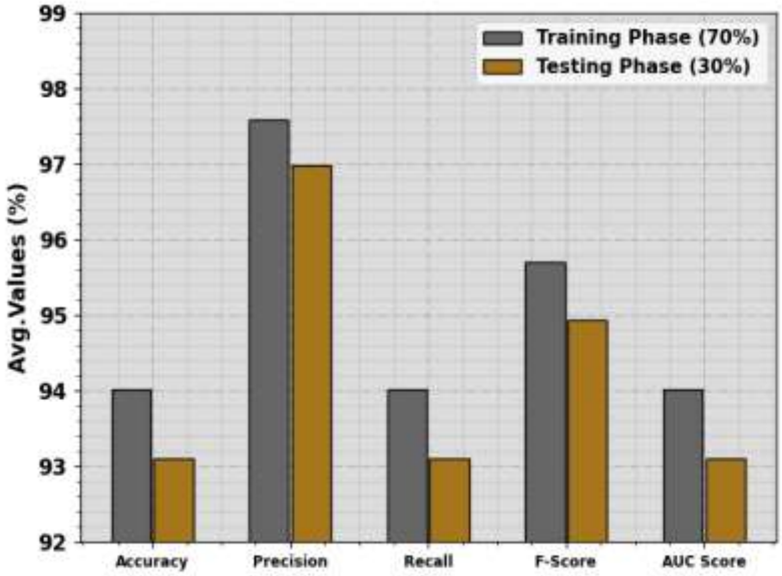

Figure 5: Average of IDA-HGOAML system on 70:30 of TRPH/TSPH

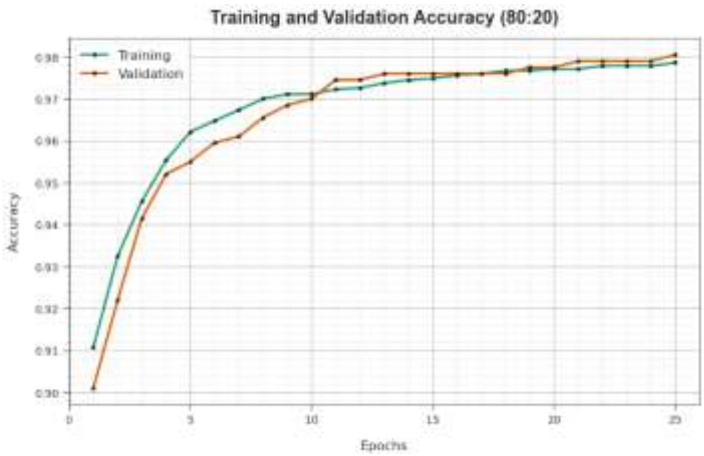

Figure 6: $Accu_y$ curve of IDA-HGOAML approach at 80:20 of TRPH/TSPH

The TR and TS $accu_y$ curves evaluate the performance of IDA-HGOAML model at 80:20 of TRPH/TSPH, as demonstrated in Fig. 6. The TR and TS $accu_y$ curves illustrate the outcomes of IDA-HGOAML method on different epochs. The figure provides relevant details based on the learning task and generalizability of the IDA-HGOAML approach. The TR and TS $accu_y$ curves are improved with increasing epoch count. It is experimental that the IDA-HGOAML method accomplishes maximum testing accuracy which can identify the patterns in the TR and TS datasets.

The TR and TS loss values of IDA-HGOAML model at 80:20 of TRPH/TSPH over epochs is depicted in Fig. 7. The TR loss shows that the model loss is lesser over epoch counts. At first, the loss value is decreased as the model adapts the weight to minimize the predictive errors on the TR and TS datasets. The loss curves illustrate the how well the model fits the training datasets. The TR and TS loss is declined progressively which signifies that the IDA-HGOAML method effectively learns the patterns shown in the TR and TS datasets. The IDA-HGOAML system finetune the parameter for minimizing the discrepancy between the original training labels and the predictions.

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

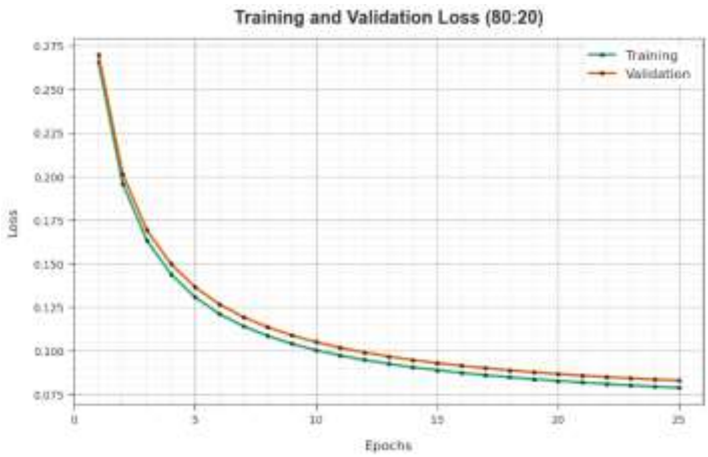

Figure 7: Loss curve of IDA-HGOAML method at 80:20 of TRPH/TSPH

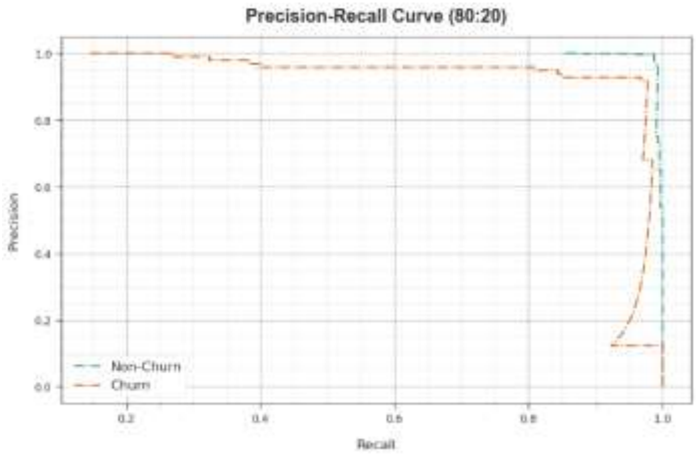

Figure 8: PR curve of IDA-HGOAML method at 80:20 of TRPH/TSPH

The PR outcome of the IDA-HGOAML algorithm at 80:20 of TRPH/TSPH is shown by plotting precision against recall as demonstrated in Fig. 8. The outcomes inferred that the IDA-HGOAML method obtains maximum PR outcomes under each class label. The outcomes shows that the model learns to recognize dissimilar classes. The IDA-HGOAML method obtains enhanced solution in detecting positive instance with least false positive.

The ROC curves presented by the IDA-HGOAML method at 80:20 of the TRPH/TSPH are demonstrated in Fig. 9. The outcomes imply the tradeoffs between the TPR and FPR rates with different classifier thresholds and dissimilar epoch counts. It describes the accurate prediction outcomes of the IDA-HGOAML system on the classifier of different classes.

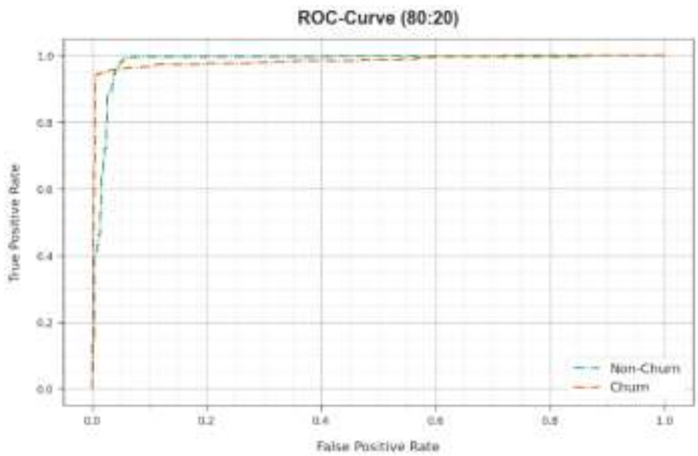

Figure 9: ROC curve of IDA-HGOAML method at 80:20 of TRPH/TSPH

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

Table 4 reports comprehensive comparative outcomes of the IDA-HGOAML system [25]. Fig. 10 inspects an extensive u_y, $prec_n$, and $reca_l$ outcomes of the IDA-HGOAML method with existing models. The outcome demonstrates that the DT method reaches worse results whereas the LR, ISMOTE-OWELM, SVM, SGD-NN, and RMSProp-NN models obtain slightly better values of ccu_y, $prec_n$, and $reca_l$. Although the AIJOA-CPDE model reaches considerable $accu_y$, $prec_n$, and $reca_l$ values of 91.41%, 97.63%, and 91.41%, the IDA-HGOAML technique confirms its better performance with $accu_y$, $prec_n$, and $reca_l$ of 94.15%, 97.91%, and 94.15% correspondingly.

Table 4: Comparative outcomes of IDA-HGOAML method with existing models

Methods	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	AUC_{Score}
IDA-HGOAML	94.15	97.91	94.15	95.92	94.15
AIJOA-CPDE	91.41	97.63	91.41	94.20	91.41
LR	80.65	79.44	80.56	79.17	82.30
DT	76.78	56.90	75.81	65.09	78.37
ISMOTE-OWELM	90.59	91.78	89.51	89.77	89.96
SVM	84.41	84.66	84.11	85.71	84.10
SGD-NN	84.54	86.22	85.92	84.43	84.91
RMSProp-NN	87.48	85.29	85.30	85.20	86.38

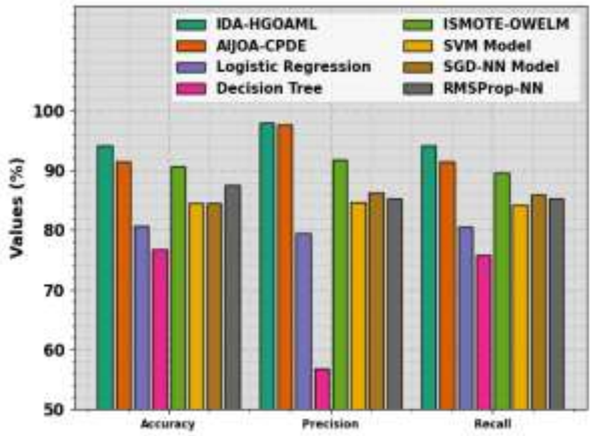

Figure 10: $Accu_y$, $prec_n$, and $reca_l$ outcome of IDA-HGOAML technique with recent methods

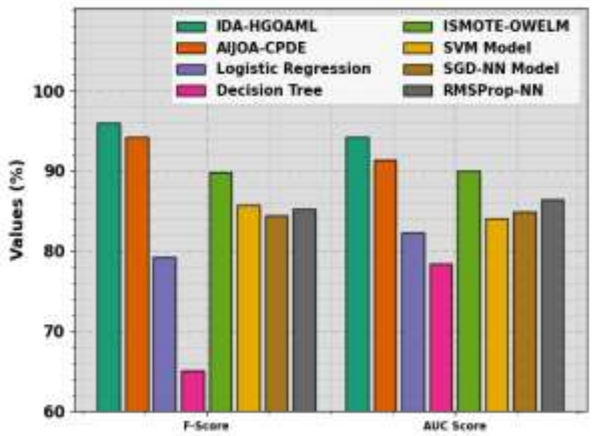

Figure 11: F_{score} and AUC_{score} outcome of IDA-HGOAML method with existing techniques

Fig. 11 examines extensive F_{score} and AUC_{score} outcomes of the IDA-HGOAML technique with existing techniques. The outcomes illustrated that the DT approach reaches worse outcomes whereas the LR, ISMOTE-OWELM, SVM, SGD-NN, and RMSProp-NN techniques attain slightly enhanced values of F_{score} and AUC_{score}. Even though the AIJOA-CPDE method obtains considerable F_{score} and AUC_{score} values of 94.20% and 91.41%,

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

the IDA-HGOAML technique confirms its better performance with F_{score} and AUC_{score} of 95.92% and 94.15% correspondingly. These outcomes confirmed the superior performance of the IDA-HGOAML algorithm on the CCP technique.

5. Conclusion

In this study, a new IDA-HGOAML model was introduced for the prediction of customer churns. The main objective of IDA-HGOAML method focuses on the prediction and classification of customer churns and non-churns. To do so, the IDA-HGOAML technique incorporates different processes namely data pre-processing using Z-score normalization, EOA-based FS, CAE classification, and HGOA-based parameter tuning. In this study, the CAE model is employed for the effectual recognition of customers into churners and non-churners. Lastly, the HGOA is exploited for the optimum hyperparameter selection of CAE model, thereby improving the prediction results. A detailed set of experiments were carried out to validate the higher efficiency of the IDA-HGOAML method. The extensive outcomes indicated the better prediction results of the IDA-HGOAML technique over existing methods under different measures.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Wahul, R.M., Kale, A.P. and Kota, P.N., 2023. An Ensemble Learning Approach to Enhance Customer Churn Prediction in Telecom Industry. International Journal of Intelligent Systems and Applications in Engineering, 11(9s), pp.258-266.
[2] Fujo, S.W., Subramanian, S. and Khder, M.A., 2022. Customer churn prediction in the telecommunication industry using deep learning. Information Sciences Letters, 11(1), p.24.
[3] Larasati, A., Ramadhanti, D., Chen, Y.W. and Muid, A., 2021, October. Optimizing Deep Learning ANN Model to Predict Customer Churn. In 2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE) (pp. 1-5). IEEE.
[4] Thorat, M.A.S. and Sonawane, V.R., 2023. CUSTOMER CHURN PREDICTION IN TELECOMMUNICATION INDUSTRY USING DEEP LEARNING. Journal of Data Acquisition and Processing, 38(3), p.1417.
[5] Sudharsan, R. and Ganesh, E.N., 2022. A Swish RNN-based customer churn prediction for the telecom industry with a novel feature selection strategy. Connection Science, 34(1), pp.1855-1876.
[6] Motevali, M., 2023. Predicting Customer Churn Based on Deep Learning, Neural Networks and Logistic Regression. Mathematical Statistician and Engineering Applications, 72(1), pp.2180-2190.
[7] Almufadi, N. and Qamar, A.M., 2022. Deep Convolutional Neural Network Based Churn Prediction for Telecommunication Industry. Comput. Syst. Sci. Eng., 43(3), pp.1255-1270.
[8] Kiruthika, J.K., Neshamoney, B.S., Kumar, L.M. and Mugunthan, V., 2021, July. Churn Prediction in Telecom sector using Deep Neural Network with Flask Application. In IOP Conference Series: Materials Science and Engineering (Vol. 1166, No. 1, p. 012060). IOP Publishing.
[9] Wahul, R.M., Kale, A.P. and Kota, P.N., 2023. An Ensemble Learning Approach to Enhance Customer Churn Prediction in Telecom Industry. International Journal of Intelligent Systems and Applications in Engineering, 11(9s), pp.258-266.
[10] Zaky, A., Ouf, S. and Roushdy, M., 2022, March. Predicting banking customer churn based on artificial neural network. In 2022 5th International Conference on Computing and Informatics (ICCI) (pp. 132-139). IEEE.
[11] Jajam, N., Challa, N.P., Prasanna, K.S. and Ch, V.S.D., 2023. Arithmetic Optimization with Ensemble Deep Learning SBLSTM-RNN-IGSA model for Customer Churn Prediction. IEEE Access.
[12] Amatare, S.A. and Ojo, A.K., 2020. Predicting customer churn in telecommunication industry using convolutional neural network model. IOSR Journal of Computer Engineering (IOSR-JCE), 22(3), pp.54-59.
[13] Gabhane, M.D. and Aslam Suriya, D.S., 2022. Churn Prediction in Telecommunication Business using CNN and ANN. Journal of Positive School Psychology, pp.4672-4680.
[14] Abualkishik, A.Z., Almajed, R. and Thompson, W., 2023. Intelligent Model for Customer Churn Prediction using Deep Learning Optimization Algorithms. Journal of Intelligent Systems & Internet of Things, 8(1).
[15] Jeyakarthic, M. and Venkatesh, S., 2020. An effective customer churn prediction model using adaptive gain with back propagation neural network in cloud computing environment. Journal of Research on the Lepidoptera, 51(1), pp.386-399.
[16] Pekel Ozmen, E. and Ozcan, T., 2022. A novel deep learning model based on convolutional neural networks for employee churn prediction. Journal of Forecasting, 41(3), pp.539-550.

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024

[17] Garimella, B., Prasad, G.V.S.N.R.V. and Prasad, M.K., 2021. Churn prediction using optimized deep learning classifier on huge telecom data. Journal of Ambient Intelligence and Humanized Computing, pp.1-22.

[18] Amin, A., Adnan, A. and Anwar, S., 2023. An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes. Applied Soft Computing, 137, p.110103.

[19] Hussen, N., Elghamrawy, S.M., Salem, M. and El-Desouky, A.I., 2023. A Fully Streaming Big Data Framework for Cyber Security based on Optimized Deep Learning Algorithm. IEEE Access.

[20] Alham, M.H., Gad, M.F. and Ibrahim, D.K., 2023. Potential of wind energy and economic assessment in Egypt considering optimal hub height by equilibrium optimizer. Ain Shams Engineering Journal, 14(1), p.101816.

[21] Mafarja, M., Thaher, T., Al-Betar, M.A., Too, J., Awadallah, M.A., Abu Doush, I. and Turabieh, H., 2023. Classification framework for faulty-software using enhanced exploratory whale optimizer-based feature selection scheme and random forest ensemble learning. Applied Intelligence, pp.1-43.

[22] Seyfioğlu, M.S., Özbayoğlu, A.M. and Gürbüz, S.Z., 2018. Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities. IEEE Transactions on Aerospace and Electronic Systems, 54(4), pp.1709-1723.

[23] Zhang, Y., Li, T., Li, Y. and Wang, G., 2023. Parameter Identification of Pilot Model and Stability Analysis of Human-in-Loop Image Seeker. Aerospace, 10(9), p.806.

[24] https://www.kaggle.com/datasets/arashnic/telecom-churn-dataset

[25] Abdullaev, I., Prodanova, N., Ahmed, M.A., Lydia, E.L., Shrestha, B., Joshi, G.P. and Cho, W., 2023. Leveraging metaheuristics with artificial intelligence for customer churn prediction in telecom industries. Electronic Research Archive, 31(8), pp.4443-4458.

Doi: https://doi.org/10.54216/FPA.140213
Received: July 16, 2023 Revised: November 22, 2023 Accepted: January 19, 2024