Intelligent System for Customer Churn Prediction using Dipper Throat Optimization with Deep Learning on Telecom Industries

Sergey Bakhvalov[1], Eduard Osadchy[2], Irina Bogdanova[3], Rustem Shichiyakh[4], E. Laxmi Lydia[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 of Department of Economics and Management, Head of the Faculty of Economic and Legal Sciences, Kazan Federal University, Elabuga Institute of KFU, Elabuga, Russia:
[3]Candidate of Economic Sciences, Associate Professor of Department of Economics and Management, Don State Technical University, Rostov-on-don, Russia;
[4]Candidate of Economic Sciences, Associate Professor of Department of Management, Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia;
[5]Department of Computer Science and Engineering, GMR Institute of Technology, Andhra Pradesh, Rajam 532127, India;
Emails: bakhvalov.s.yu@yandex.ru; eosadchy@mail.ru; bogdanova.i.v@inbox.ru; shichiyakh.r.a@mail.ru; elaxmi2002@yahoo.com

Abstract

Intelligent System for Customer Churn Prediction (CCP) relates to a system or application that controls advanced artificial intelligence (AI), data analysis, and machine learning (ML) methods for anticipating and predicting customer churn in business or service. CCP approach utilizes various data sources comprising customer behavior and historical data, to create predictive method able of categorizing customers who are potential to leave or stop their engagement. By employing intelligent method, this system supports businesses in proactively addressing customer retention and executing manners to decrease churn, ultimately enhancing revenue retention and customer satisfaction. It connects wide data sources, comprising customer behavior and historical information, to progress difficult methods that can identify customers at risk of leaving or discontinuing their service or subscription. By leveraging deep learning (DL) method, this intelligent system enhances the efficiency and accuracy of customer churn prediction, allowing businesses to take proactive measures to maintain customers, maintain revenue, and develop customer satisfaction. This article presents an Intelligent System for Customer Churn Prediction using Dipper Throat Optimization with Deep Learning (ISCCP-DTODL) methodology in Telecom Industries. The purpose of the ISCCP-DTODL system focuses on the design of intelligent systems for the effective prediction of customer churners and non-churners. To accomplish this, the ISCCP-DTODL system performs Z-score data normalization to preprocess the data. For feature selection and to reduce high dimensionality of features, the ISCCP-DTODL technique uses DTO algorithm. Besides, the ISCCP-DTODL technique makes use of hybrid CNN-BiLSTM model for churn prediction. At last, jellyfish optimization (JFO) based hyperparameter tuning approach can be employed to pick hyperparameters connected to CNN-BiLSTM technique. To display enhanced performance of ISCCP-DTODL technique, a widespread set of simulations was performed. The extensive results stated that ISCCP-DTODL model illustrates improved results than its current techniques in terms of dissimilar measures.

Keywords: Customer Churn Prediction; Deep Learning; Dipper Throat Optimization; Parameter tuning; Jellyfish Optimization

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

1. Introduction

Customer churn (CC) is described as when customers discontinue business performance with a company [1]. This occurrence has directly affected insurance, banks, telecom businesses, and video game companies. Recently, telecommunication industries endured large modifications like raised competition, technological development, and new services [2]. Consequently, forecasting customer churn in the telecommunication field is a crucial factor in industrial activities for protecting their reliable customers. The telecom field has been major significant industry in developed nations in previous two decades [3]. Data mining performs an important function for predicting and analyzing in the telecom sector because of accessibility of enormous information. The fundamental application field is to execute churner prediction for saving customer maintenance and to generate a higher-profit rate [4].

CC belongs to loss of a customer in interest of a participant, considering the relationship end [5]. Customer churn prediction (CCP) permits any to detect the causes for ending the correlation as well as collecting an approach that can be reduce the churn rate, and improve profits. Therefore, expecting a customer's purpose to finale a relationship can be effectual in telecommunication industries and regarded viable benefits [6]. In recent years, it has a remarkable increase to implement ML methods for CCP in various sectors. By utilizing these methods with data acquired from customers telecommunication operator's produces consistent methods that could be represented connection and correlation, overcoming customer churn problems in telecommunications [7]. Present research work demonstrate that an effective churn prediction method must proficiently utilize massive volume of past data for enriched churner's detection. But, there are numerous limits in present methods because of it could not be probable to proficiently execute churn prediction with higher accuracy. A massive data volume has bene produced at telecom sector, which comprises missing values [8]. Prediction under this kind of information leads to incorrect or poorer outputs for predictive methods in this study. Data preprocessing has been now implemented for determining this problem and missing values imputation could be executed employing ML technique that accelerates higher effectiveness and prediction or classification accuracy [9]. Feature selection (FS) is also performed in this research work; but a few significant and data rich features have been disregarded in development of model. Additionally, statistical approaches could be employed for developing model that causes lower performance of prediction. Also, benchmark databases have not utilized for model assessment in this study that gives rise to reduce the representation of real data image. Reasonable comparison among various techniques are also implemented with no standard database [10]. An intelligent system is exploited for solving the present problems and offering higher-accuracy churn prediction.

This article introduces an Intelligent System for Customer Churn Prediction using Dipper Throat Optimization with Deep Learning (ISCCP-DTODL) technique in Telecom Industries. The goal of the ISCCP-DTODL technique focuses on the design of intelligent systems for the effectual prediction of customer churners and non-churners. To accomplish this, the ISCCP-DTODL technique performs Z-score data normalization to preprocess the data. For FS and to reduce high dimensionality of features, the ISCCP-DTODL technique uses DTO algorithm. Besides, the ISCCP-DTODL technique makes use of hybrid CNN-BiLSTM model for churn prediction. At last, jellyfish optimization (JFO) based hyperparameter tuning approach can be used to pick hyperparameters connected to CNN-BiLSTM technique. To exhibit enhanced performance of ISCCP-DTODL methodology, a widespread set of simulations was implemented.

2. Related works

Zhao et al. [11] developed a customer performance detection technique that integrates unsupervised and supervised classification models. Initially, a hybrid technique of K-means clustering and entropy model and client portrait examination used in order to divide consumers. Next, division outcomes are then united into presented multihead self-attention depend nested LSTM classification algorithm. Third, developed architecture applied to real case. Nalatissifa and Pardede [12] developed a model of DNN. This research employs 3 variants of hidden layers (HLs) and 2 variants of activation functions such as Sigmoid and rectified linear unit (ReLu), then employs 5 variants of optimizers. Sudharsan and Ganesh [13] develops a new design in order to forecast customer churn (CC) via DL technique such as Swish Recurrent Neural Network (S-RNN).

Xu et al. [14] creates a telecom customer churn forecast method with aid of backpropagation neural network (BPNN) technique as well as uses MapReduce programming structure on Hadoop. By employing information of telecom company, this study examines harm of telecom consumers in great information atmosphere. The study displays exactness of telecom CC estimate approach in BPNN. In [15], a new structure conveyed employing reformatted RNN (R-RNN) in combination with Elephant herding optimizer (EHO) technique to estimate customer turnover. To categorize CC and normal customers, RRNN is reformed. This amended EHO efficiently raises exact RNN parameters. The distance was upgraded by EHO employing a clan operator. Pustokhina et al. [16] projects a novel improved SMOTE by optimum weighted extreme machine learning (OWELM) termed as

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

ISMOTE-OWELM techniques. After standardization, ISMOTE employed in order to grip extreme dataset of rain optimization algorithm (ROA) functional to regulate optimum sampling rate. Finally, WELM method was applied.

In [17], a hybrid technique depend on data mining technique projected in order to analyse features of CC. Originally, FS node was utilized to classify features. Then, Bayesian network and C5.0 DT are employed for recognition. These are data mining models and terms that can aid in predicting. At last, projected technique executed in chain store trade as a case research. Liu et al. [18] developed BiLSTM-CNN model combined with RNN and CNN approaches. This article employs bank data to associate attention BLSTM-CNN (AttnBLSTM-CNN) model with BiLSTM-CNN methodology.

3. The Proposed Model

In this article, we presented novel ISCCP-DTODL model on Telecom Industries. The goal of the ISCCP-DTODL technique focuses on the design of intelligent systems for the effectual prediction of customer churners and non-churners. To accomplish this, the ISCCP-DTODL technique performs such as Z-score data normalization, DTO-based FS, CNN-BiLSTM-based classification, and JFO-based hyperparameter tuning. Fig. 1 demonstrates overall procedure of ISCCP-DTODL system.

A. Data Normalization

In this work, the ISCCP-DTODL technique performs Z-score data normalization to preprocess the data. Z-score data normalization, otherwise called standardization, is a statistical approach used to convert the data into uniform distribution with mean of 0 and standard deviation (SD) of 1 [19]. This technique includes subtracting the mean of data from all the data points and later dividing the results by the SD. Z-score normalization is more commonly used in statistical modeling, data analysis, and machine learning to ensure that data with dissimilar scales and units are analyzed and compared on a regular basis, which makes it a powerful tool to improve the performance of various analytical approaches and achieve data consistency.

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

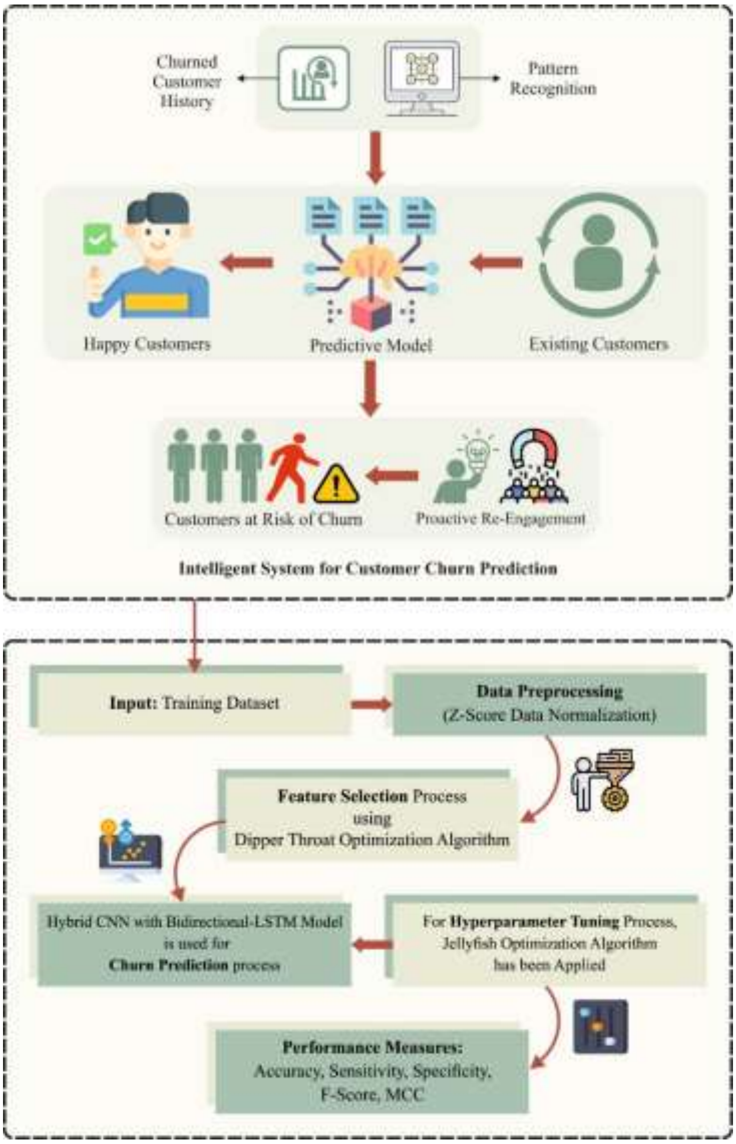


Figure 1: Overall process of ISCCP-DTODL algorithm

B. Feature selection using DTO

For feature selection, ISCCP-DTODL model uses DTO algorithm. Dipper-throated passerines are rare [20]. They hunt, swim, and dive well. The DTO technique considers birds fly and swim to search for food, with N_{fs} representing amount of birds. Bird location is N_{fs} and velocity is BV as follows:

$$BP = \begin{bmatrix} BP_{1,1} & BP_{1,2} & BP_{1,3} & \dots & BP_{1,d} \\ BP_{2,1} & BP_{2,2} & BP_{2,3} & \dots & BP_{2,d} \\ BP_{3,1} & BP_{3,2} & BP_{3,3} & \dots & BP_{3,d} \\ \dots & \dots & \dots & \dots & \dots \\ BP_{n,1} & BP_{n,2} & BP_{n,3} & \dots & BP_{n,d} \end{bmatrix} \quad (1)$$

In Eq. (1), the i^{th} bird location in j^{th} dimension is represented as $BP_{i,j}$.

$$BV = \begin{bmatrix} BV_{1,1} & BV_{1,2} & BV_{1,3} & \dots & BV_{1,d} \\ BV_{2,1} & BV_{2,2} & BV_{2,3} & \dots & BV_{2,d} \\ BV_{3,1} & BV_{3,2} & BV_{3,3} & \dots & BV_{3,d} \\ \dots & \dots & \dots & \dots & \dots \\ BV_{n,1} & BV_{n,2} & BV_{n,3} & \dots & BV_{n,d} \end{bmatrix} \quad (2)$$

In Eq. (2), the i^{th}velocity of birds at j^{th} dimension is $BV_{i,j}$. The value of objective function, f_n, is described by the following expression:

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

$$f = \begin{bmatrix} f_1(BP_{1,1}, BP_{1,2}, BP_{1,3}, \dots, BP_{1,d}) \\ f_2(BP_{2,1}, BP_{2,2}, BP_{1,3}, \dots, BP_{2,d}) \\ f_3(BP_{3,1}, BP_{3,2}, BP_{3,3}, \dots, BP_{3,d}) \\ \dots \\ f_n(BP_{n,1}, BP_{n,2}, BP_{n,3}, \dots, BP_{n,d}) \end{bmatrix} \quad (3)$$

Next, the values of objective function arranged from lower to higher in ascending order. The initial optimum solution is denoted as BP_{best}. BP_{Gbest} is best possible performance. It is predicted that the remaining response have pertained to classic bird. BP_{nd} is follower bird. The position of swimming bird shifts as it swims:

$$BP_{nd}(t+1) = BP_{best}(t) - C_1.|C_2.BP_{best}(t) - BP_{nd}(t)| \quad (4)$$

In Eq. (4), parameters C_1 and C_2 are defined as $C_1 = 2c.r_1 - c$ and $C_2 = 2r_1$ for $c = 2$, upgraded from two to zero, r_1 is randomly upgraded in zero and one; and $T_{\max}$ denotes maximal iterations.

The position of flying birds can upgraded by following expression:

$$BP_{nd}(t+1) = BP_{nd}(t) + BV(t+1) \quad (5)$$

The flying velocity bird is changed as:

$$BV(t+1) = C_3BV(t) + C_4r_2\big(BP_{best}(t) - BP_{nd}(t)\big) + C_5r_2\big(BP_{Gbest} - BP(t)\big) \quad (6)$$

In Eq. (6), C_3 denotes weight value, C_4 and C_5 are constants. r_2 is randomly updated within [0,1]. Fig. 2 illustrates steps involved in DTO.

Algorithm 1: Pseudocode of DTO Algorithm

Initialize the position of bird as $P_i(i = 1,2, n)$, velocity as $BV_i(i = 1,2, \dots, n)$, iteration $T_{\max}$, objective function f_n, other DTO parameters, $t = 1$
 	
Compute f_n for every bird BP_i
Find the optimum bird BP_{best}
while $t \leq T_{\max}$ do
  for $(i = 1{:}\, i < n + 1)$ do
    if $(R < 0.5)$ then
      Upgrade swimming position of the bird
      $BP_{nd}(t+1) = BP_{best}(t) - C_1.|C_2.BP_{best}(t) - BP_{nd}(t)|$
    else
      Upgrade velocity of flying bird as
      $BV(t+1) = C_3BV(t) + C_4r_2\big(BP_{best}(t) - BP_{nd}(t)\big) + C_5r_2\big(BP_{Gbest} - Bp(t)\big)$
      Update location of flying bird
      $BP_{nd}(t+1) = BP_{nd}(t) + BV(t+1)$
    End if
  end for
  Update f_n for every bird BP_i
  Update parameter, $t = t + 1$
  Update the optimum bird BP_{best}
  Set $BP_{Gbest} = BP_{best}$
end while
Return BP_{Gbest}

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

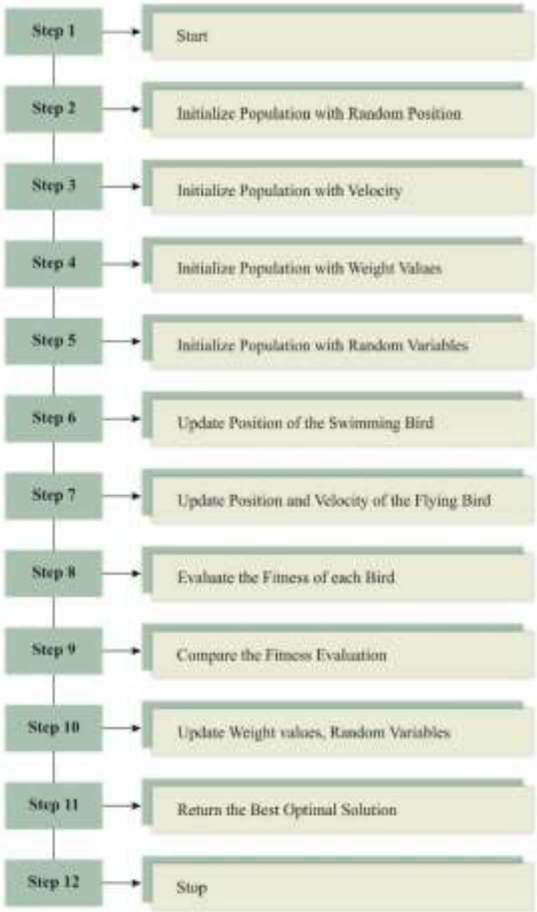


Figure 2: Steps involved in DTO

The fitness function considers classifier outcome and amount of attributes particular. It minimizes set size of attributes selected and maximizes the classification outcome. Thus, fitness function is used to measure individual solutions as follows.

$$Fitness = \alpha * ErrorRate + (1 - \alpha) * \frac{\#SF}{\#All\_F} \tag{7}$$

In Eq. (7), $ErrorRate$ signifies the classifier error rate using the attributes selected. $ErrorRate$ can be estimated as the percentage of incorrect classified to the count of classifiers made, within [0,1]. ($ErrorRate$ is complement of classifier accuracy), number of features selected is $\#SF$ and overall amount of features in the original dataset is $\#All\_F$. α is used to manage importance of classifier quality and subset length which fixed as 0.9.

C. Classification using CNN-BiLSTM

At this stage, the ISCCP-DTODL technique makes use of hybrid CNN-BiLSTM model for churn prediction. The developed technique consists of dual different blocks. The CNN block includes a flatten operator, a CNN layer, and a pooling layer [21]. The resultant of CNN block is utilized as an input by Bi-LSTM block. The aim of pooling and convolutional layers is to filter incoming data for extracting relevant data from the matrix. The convolutional operation can be done by the convolution layers between input data and small matrices named filters or kernels. The convolutional function is formulated below. Consider n and m indicate indexes of rows and columns of resultant matrix (R), H as an input matrix, and I refers to kernel matrix.

$$R[m, n] = (H \cdot I)[m, n] \tag{8}$$

$$= \sum_j \sum_k I\,[j, k] \cdot H[m - j, n - k]$$

The convolution layer exploits $ReLU$ activation function more generally employed in CNN technique. The major benefit is that it doesn't activate all the neurons simultaneously since it transforms each negative value into 0. Consequently, $ReLU$ has a large computation efficacy:

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

$$f(x) = x^{+} = \max(0, x) \quad (9)$$

BiLSTM block is second module encompassed by a dense layer, Bi-LSTM network, and a dropout layer. First, we describe a (unidirectional) LSTM model to understand the architecture and performance of BiLSTM technique. The LSTM is a kind of RNN. Classical RNNs do not have a predetermined structure of the layer. All the LSTM units are comprised of three gates (forget, input, and output gates) and memory unit which regulates the data flow by deciding what data to forget and retain.

In forget gate, a sigmoid function can be used to data added in prior HL (h-1) and the existing input (X_t). The $f_t$function returns the value amid 0 and 1 that shows amount of data that would be retained.

The input gate takes data from existing input and from prior HL and passes them through second sigmoid function, which transforms this data into values within [0,1]. Also, the same data passed over a $tanh$ function, enables regulating the network, which returns the value within [-1,1]. Next, sigmoid output (i_t) is multiplied with the $tanh$ output ($\tilde{C}_t$) to define what data to be kept.

Firstly, the prior cell state (C_{t-1}) is multiplied by forgotten output. Next, output of input gate is added that upgrades cell state. The new cell state (C_t) is outcome of these two operations.

Lastly, we have output gate. It defines value of next HL. Initially, data from the existing input and from the prior HL passed over third sigmoid function. Next, novel cell state passed over a $tanh$ function. This output was point-by-point multiplied. The output gate (0_t) is novel HL (h_t). Thus, new HL and cell state will be passed to next timestep:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_C) \quad (10)$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \cdot tanh(C_t)$$

Where tanh denotes the hyperbolic tangent function, σ indicates sigmoid function, h_t signifies the HL at t time, W_x represents a weight matrix, x_t characterizes the input dataset at t time and b_x is a bias vector.

A Bi-LSTM network has forward and backward LSTM networks. The forward LSTM exploits an input as a series of values ranging from $t-k$ to t whereas the backward LSTM exploits an as input a series ranging from t to $t-k$. The output of backward ($\overleftarrow{h}$) and forward ($\overrightarrow{h}$) networks are calculated by single LSTM.

$$Y_t = \sigma\left(\overrightarrow{h_t}, \overleftarrow{h_t}\right) \quad (11)$$

The sigmoid function combines output of single LSTM network. Bi-LSTM uses the dropout model to regulate overfitting. Lastly, an FC layer returns predictive output.

D. Hyperparameter Tuning JFO

At last, the JFO approach can be employed to pick hyperparameters connected to CNN-BiLSTM technique. JFO algorithm demonstrated on swimming performance of jellyfish (JF) [22]. In this technique, clarification for every problem is a JF that appears mainly for food or an optimum result. The sea's waves offer food that might appeal to JF. Eq. (12) demonstrates how can be employed to describe way of ocean is present:

$$\overrightarrow{trend} = \frac{1}{nPop} . \sum \overrightarrow{trend}_i = \frac{1}{nPop} \sum (X^* - e_c X_i) \quad (12)$$

Here, e_c symbolize absorption factor Eq. (13) generated by covering Eq. (12).

$$\overrightarrow{trend} = X^* - \frac{\Sigma e_c X_i}{nPop} = X^* - e_c\mu \quad (13)$$

The finest JF, X^*denotes average JF populace, μ, utilized in this evaluation. Then $df = e_c\mu$ expected, this equation is written in common method as Eq. (14):

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

$$\overrightarrow{trend} = X^* - \frac{\Sigma e_c X_i}{nPop} = X^* - df \tag{14}$$

Eqs. (15) and (16) permit to reflect supply of JF from arbitrary to normal.

$$df = \beta \times \sigma \times rand^f(0,1) \tag{15}$$

$$\sigma = rand^f(0,1) \times \mu \tag{16}$$

The JF distribution's ordinary deviation index is signified in these contacts by representation σ. The normal distribution of JF dispersed near mean point.

The movement procedure of each JF below affects of JF community as well as aquatic force of sea.

Eqs. (17) and (18) were employed to change equations df and e_c, correspondingly:

$$df = \beta \times rand(0,1) \times \mu \tag{17}$$

$$e_c = \beta \times rand(0,1) \tag{18}$$

Eq. (14) existing in Eq. (19) after being modified rely on Eq. (17):

$$\overrightarrow{trend} = X^* - \beta \times rand(0,1) \times \mu \tag{19}$$

They are struggling with JF aquatic waves as revealed by Eq. (20):

$$X_i(t+1) = X_i(t) + rand(0,1) \times \overrightarrow{trend} \tag{20}$$

It is highly probable to enlarge Eq. (20) to Eq. (21):

$$X_i(t+1) = X_i(t) + rand(0,1) \times (X^* - \beta \times rand(0,1) \times \mu) \tag{21}$$

In this equation, β is frequently equivalent to 3 and greater than zero. JF also transfers in collections and often substitutes among dual passive and vigorous actions. When they are lazy, they will stare more closely at their environments. Eq. (22) utilized to pretend inactive signal:

$$X_i(t+1) = X_i(t) + \gamma . rand(0,1) \times (U_b - L_b) \tag{22}$$

In Eq. (22), continuous motion is an optimistic number and naturally set at 0.1, signified by character γ. Each dimension of U means upper value and L denotes lower value specified by "b". The JF X_i arbitrarily selects JF X_i in active conduct, mode has dual styles. Eq. (23) used to transfer if X_i's value beats X_j's, and Eq. (24) employed in all further cases:

$$X_i(t+1) = X_i(t) + rand.\left(X_j(t) - X_i(t)\right) \tag{23}$$

$$X_i(t+1) = X_i(t) + rand.\left(X_i(t) - X_j(t)\right) \tag{24}$$

Eq. (25) is applied to transform among sea and collective actions:

$$c(t) = \left|\left(1 - \frac{t}{\text{Maxt}}\right) \times (2.rand - 1)\right| \tag{25}$$

In this equation, $Maxt$ stands for maximum iteration counter and t signifies amount of current iterations. The JF upgrade depends upon group actions if $c(t)$ is fewer than 0.5 and on waves if it is higher than 0.5 for every upgrade.

The fitness choice is main factor affecting performance of MBES technique. The hyperparameter choice manner comprises result encoded method to assess efficiency of candidate result. The MBES algorithm considers accuracy as a primary condition to develop the FF as follows.

$$Fitness = \max(P) \tag{26}$$

$$P = \frac{TP}{TP + FP} \tag{27}$$

Where TP and FP represent true and false positive values.

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

4. Performances Validation

In this section, CCP results of ISCCP-DTODL technique is examined on the churn database [23]. It contains 3333 instances with 21 features and 2 classes. The ISCCP-DTODL system can be selected a set of 11 features.

Fig. 3 shows confusion matrices made by OAFS-HDLCP system with 80:20 and 70:30 of TRA phase/TES phase. The achieved outcomes represents the efficacious identification of the churn and non-churn samples with each class.

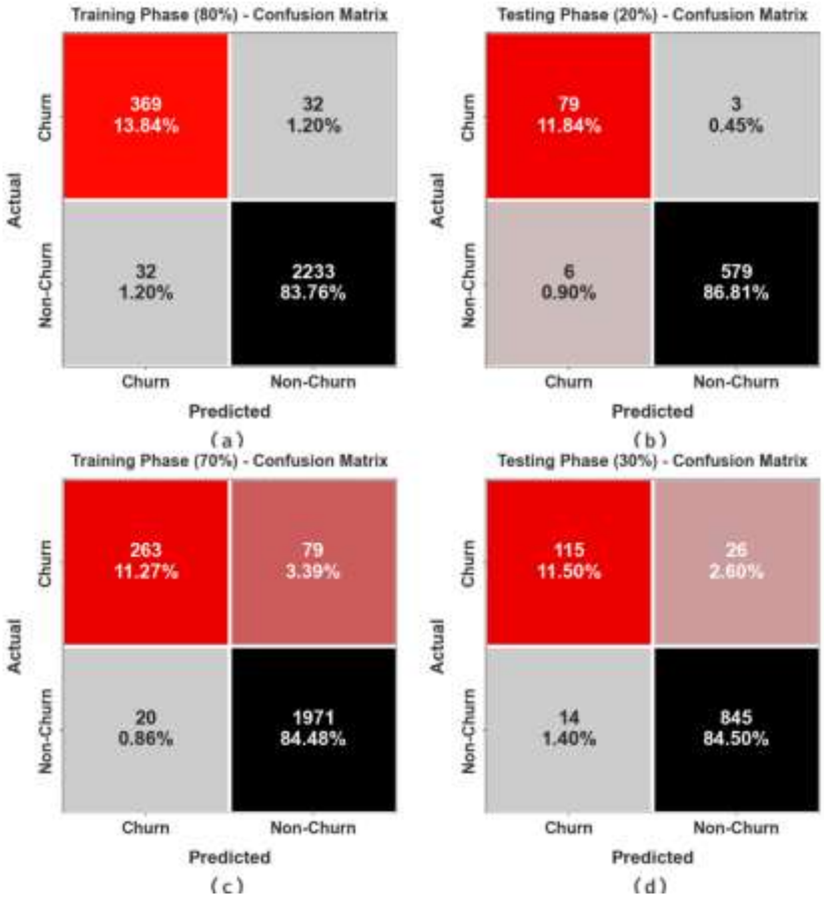


Figure 3: Confusion matrices of (a-c) TRA phase of 80% and 70% TRA phase and (b-d) TES phase of 20% and 30%

In Table 1 and Fig. 4, CCP results of ISCCP-DTODL system is examined on 80:20 of TRA phase/TES phase. The accomplished outcome shown that the ISCCP-DTODL system gets effectual identification of churn and non-churn. With 80% of TRA phase, the ISCCP-DTODL method gives average $accu_y$ of 95.30%, $prec_n$ of 95.30%, $reca_l$ of 95.30%, F_{score} of 95.30%, and AUC_{score} of 95.30%. At the same time, based on 20% of TES phase, the ISCCP-DTODL methodology gives average $accu_y$ of 97.66%, $prec_n$ of 96.21%, $reca_l$ of 97.66%, F_{score} of 96.92%, and AUC_{score} of 97.66%, respectively.

Table 1: The CCP analysis of the ISCCP-DTODL system under 80:20 of TRA phase/TES phase

Classes	$\mathbf{Accu_y}$	$\mathbf{Prec_n}$	$\mathbf{Reca_l}$	$\mathbf{F_{score}}$	$\mathbf{AUC_{score}}$
TRA phase (80%)					
Churn	92.02	92.02	92.02	92.02	95.30
Non-Churn	98.59	98.59	98.59	98.59	95.30
Average	95.30	95.30	95.30	95.30	95.30
TES phase (20%)					
Churn	96.34	92.94	96.34	94.61	97.66
Non-Churn	98.97	99.48	98.97	99.23	97.66
Average	97.66	96.21	97.66	96.92	97.66

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

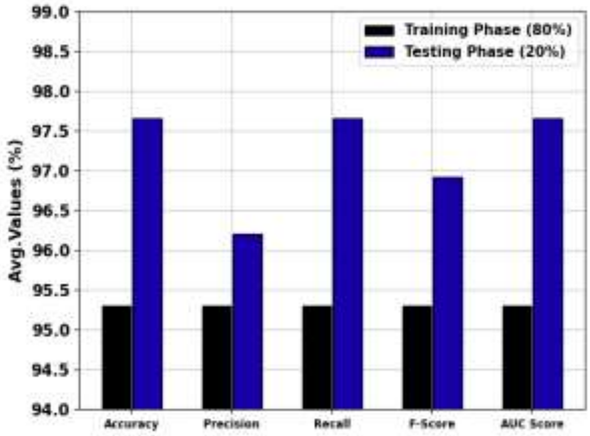


Figure 4: Average of the ISCCP-DTODL model with 80:20 of TRA phase/TES phase

In Table 2 and Fig. 5, CCP analysis of ISCCP-DTODL algorithm is described with 70:30 of TRA phase/TES phase. The attained value pointed to that ISCCP-DTODL model gets effective identification of churn and non-churn. According to 70% of TRA phase, ISCCP-DTODL methodology provides average $accu_y$ of 95.76%, $prec_n$ of 94.54%, $reca_l$ of 87.95%, F_{score} of 90.86%, and AUC_{score} of 87.95%. Concurrently, with 30% of TES phase, the ISCCP-DTODL system offers average $accu_y$ of 96.00%, $prec_n$ of 93.08%, $reca_l$ of 89.97%, F_{score} of 91.44%, and AUC_{score} of 89.97%, correspondingly.

Table 2: The CCP analysis of the ISCCP-DTODL method at 70:30 of TRA phase/TES phase

Classes	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	AUC_{score}
TRA phase (70%)					
Churn	95.76	92.93	76.90	84.16	87.95
Non-Churn	95.76	96.15	99.00	97.55	87.95
Average	95.76	94.54	87.95	90.86	87.95
TES phase (30%)					
Churn	96.00	89.15	81.56	85.19	89.97
Non-Churn	96.00	97.01	98.37	97.69	89.97
Average	96.00	93.08	89.97	91.44	89.97

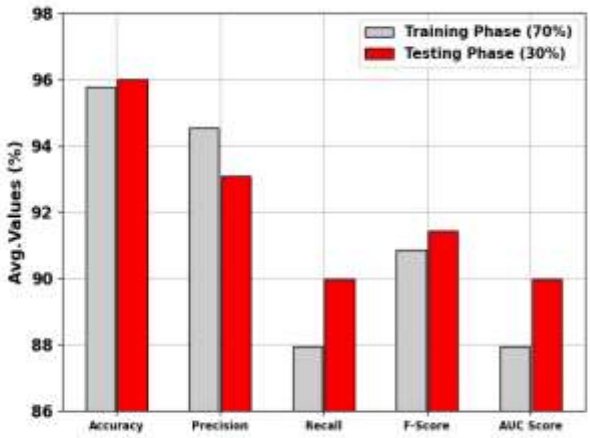


Figure 5: Average of the ISCCP-DTODL model on 70:30 of TRA phase/TES phase

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

Figure 6: $Accu_y$ curve of ISCCP-DTODL method at 80:20 of TRA phase/TES phase

To define the effectiveness of the ISCCP-DTODL system at 80:20 of TRA phase/TES phase, we have made $accu_y$ curves for TRA and TES phases, as represented in Fig. 6. Two curves gives respected insights into the model's learning progression and its proficiency for generalization. Once increasing the number of epochs, an observable growth in this TRA and TES $accu_y$ curves may be seeming. This development denoted the model's potential to greater identification patterns with respect to the databases of TRA and TES.

Fig. 7 shows an analysis of the ISCCP-DTODL model at 80:20 of TRA phase/TES phase, the loss values in the process of TRA. The minimum trends at TRA loss over epochs displays that the model repeatedly boosts their weights for lessening predictive errors with TRA and TES databases. The loss curve considers how the model could be fitted the TRA data. Especially, the TRA and TES loss dependably lower, displaying an effectual learning patterns of model existing with two datasets. Further, this reveals a difference of model to minimize variances among predictive and new TRA labels.

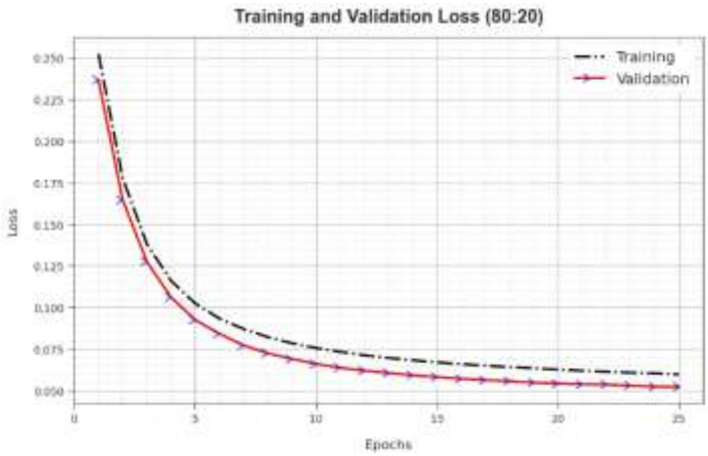
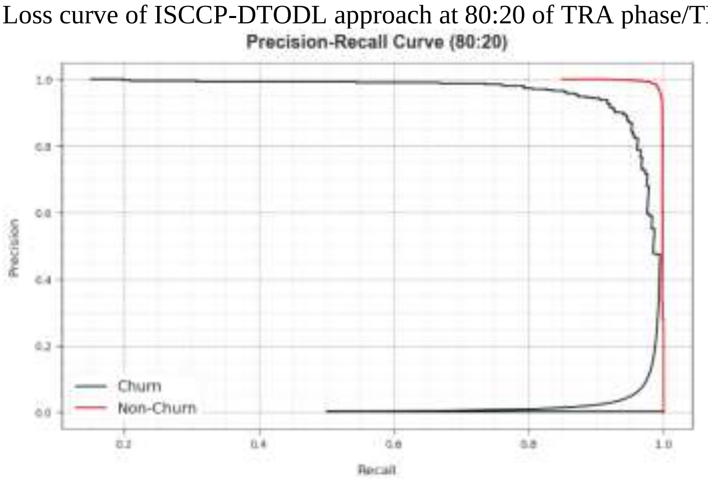


Figure 7: Loss curve of ISCCP-DTODL approach at 80:20 of TRA phase/TES phase



Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

Figure 8: Precision-recall analysis of ISCCP-DTODL approach with 80:20 of TRA phase/TES phase

In Fig. 8, the precision-recall curve of ISCCP-DTODL algorithm with with 80:20 of TRA phase/TES phase, the plots precision against recall, showing that ISCCP-DTODL model achieves better precision-recall values at each class. This graph represents the model's proficiency for identifying diverse classes, specifically exceptional to properly recognize positive samples then, diminished false positives.

Fig. 9 likewise contains ROC curves of ISCCP-DTODL methodology with 80:20 of TRA phase/TES phase that showcase the model's ability to differentiate amongst class labels. These curves offer valued insights into the trade-off among FPR and TPR at varied classification thresholds and epochs. It denotes the exact predicted performance of model within a two classes, also emphasizing its classification abilities.

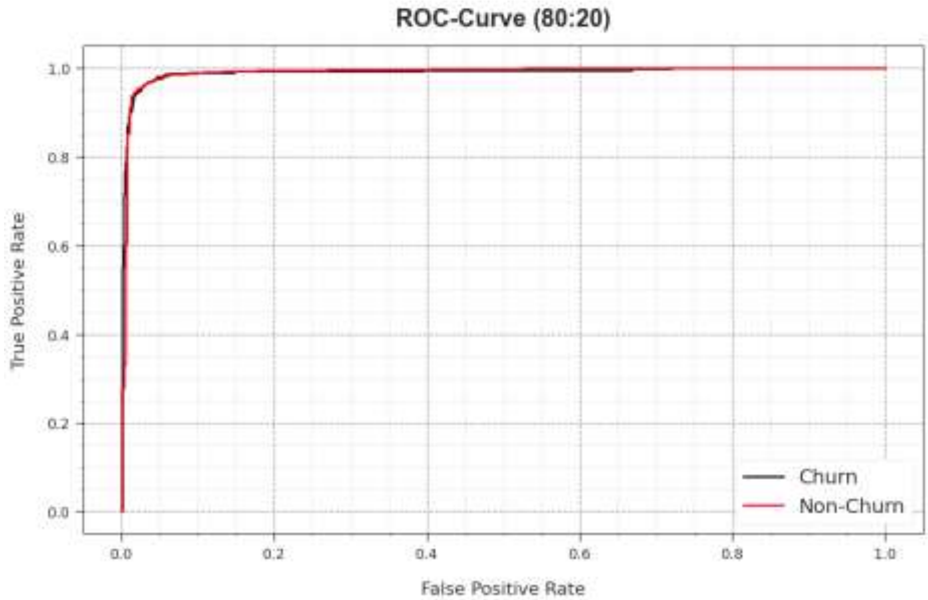


Figure 9: ROC curve of ISCCP-DTODL algorithm with 80:20 of TRA phase/TES phase

In Table 3, a comprehensive comparative outcome of ISCCP-DTODL technique for numerous measures [24]. In Fig. 10, a contrast analysis of ISCCP-DTODL system with current models with respect to $accu_y$. According to $accu_y$, the ISCCP-DTODL technique offers enhanced $accu_y$ of 97.66% while the AIJOA-CPDE, LR, DT, ISMOTE-OWELM, SVM, SGD-NN, and RMSProp-NN models obtained reduced $accu_y$ of 93.41%, 82.35%, 78.35%, 92.33%, 86.11%, 86.31%, and 88.98%, respectively.

In Fig. 11, a comparative analysis of the ISCCP-DTODL methodologies with recent models with respect to $prec_n$, and F_{score}. According to $prec_n$, the ISCCP-DTODL system offers increased $prec_n$ of 96.21% whereas the AIJOA-CPDE, LR, DT, ISMOTE-OWELM, SVM, SGD-NN, and RMSProp-NN techniques get decreased $prec_n$ of 95.63%, 81.19%, 58.65%, 93.36%, 86.32%, 87.98%, and 87.08%, correspondingly. Similarly, with F_{score}, the ISCCP-DTODL algorithm gives raised F_{score} of 96.92% but, the AIJOA-CPDE, LR, DT, ISMOTE-OWELM, SVM, SGD-NN, and RMSProp-NN models obtain diminished F_{score} of 94.32%, 80.84%, 66.68%, 91.36%, 87.22%, 85.95%, and 86.86%, correspondingly.

Table 3: Comparison analysis of ISCCP-DTODL algorithm with other approaches

Methods	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	AUC_{score}
ISCCP-DTODL	97.66	96.21	97.66	96.92	97.66
AIJOA-CPDE	93.41	95.63	94.41	94.32	94.41
Logistic Regression	82.35	81.19	82.14	80.84	84.1
Decision Tree	78.35	58.65	77.58	66.68	79.89
ISMOTE-OWELM	92.33	93.36	91.08	91.36	91.75
SVM Algorithm	86.11	86.32	85.64	87.22	85.6
SGD-NN	86.31	87.98	87.43	85.95	86.41
RMSProp-NN	88.98	87.08	86.99	86.86	88.04

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

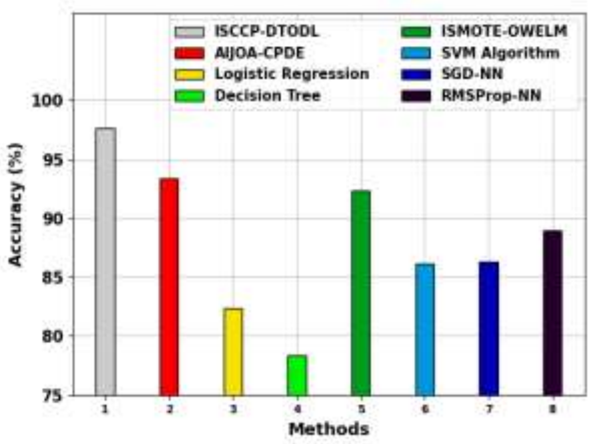


Figure 10: $Accu_y$ analysis of the ISCCP-DTODL model with other existing systems

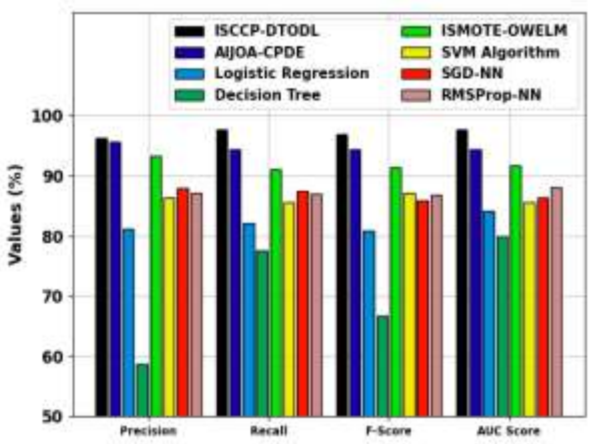


Figure 11: Comparison analysis of the ISCCP-DTODL method with other systems

Thus, ISCCP-DTODL model utilized for accurate and automated CCP process in telecom industry.

5. Conclusion

In this article, we introduce a new ISCCP-DTODL methodology on Telecom Industries. The goal of the ISCCP-DTODL technique focuses on the design of intelligent systems for the effectual prediction of customer churners and non-churners. To accomplish this, the ISCCP-DTODL technique performs such as Z-score data normalization, DTO-based FS, CNN-BiLSTM-based classification, and JFO-based parameter tuning. For FS and to reduce high dimensionality of features, the ISCCP-DTODL technique uses DTO algorithm. At last, the JFO approach can be employed to elect the hyperparameters related to the CNN-BiLSTM methodology. To reveal the enhanced solution of the ISCCP-DTODL technique, a widespread set of simulations was executed. The extensive results stated that the ISCCP-DTODL technique illustrates better results than its recent techniques in terms of dissimilar measures.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Khattak, A., Mehak, Z., Ahmad, H., Asghar, M.U., Asghar, M.Z. and Khan, A., 2023. Customer churn prediction using composite deep learning technique. Scientific Reports, 13(1), p.17294.
[2] Dalli, A., 2022. Impact of hyperparameters on Deep Learning model for customer churn prediction in telecommunication sector. Mathematical Problems in Engineering, 2022, pp.1-11.
[3] Mena, G., Coussement, K., De Bock, K.W., De Caigny, A. and Lessmann, S., 2023. Exploiting time-varying RFM measures for customer churn prediction with deep neural networks. Annals of Operations Research, pp.1-23.

Doi: https://doi.org/10.54216/FPA.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024

[4] Pondel, M., Wuczynski, M., Gryncewicz, W., Lysik, L., Hernes, M., Rot, A. and Kozina, A., 2021, July. Deep Learning for Customer Churn Prediction in E-Commerce Decision Support. In BIS (pp. 3-12).

[5] Abdulsalam, S.O., Ajao, J.F., Balogun, B.F. and Arowolo, M.O., 2022. A Churn Prediction System for Telecommunication Company Using Random Forest and Convolution Neural Network Algorithms. EAI Endorsed Transactions on Mobile Communications and Applications, 7(21).

[6] Equihua, J.P., Nordmark, H., Ali, M. and Lausen, B., 2023. Modelling customer churn for the retail industry in a deep learning based sequential framework. arXiv preprint arXiv:2304.00575.

[7] Jain, H., Khunteta, A. and Srivastava, S., 2022. Telecom churn prediction using CNN with variational autoencoder. In Smart Systems: Innovations in Computing: Proceedings of SSIC 2021 (pp. 583-600). Springer Singapore.

[8] Liu, D., Zhang, X., Shi, Y. and Li, H., 2021, August. Prediction of Railway Freight Customer Churn Based on Deep Forest. In International Conference on Intelligent Computing (pp. 479-489). Cham: Springer International Publishing.

[9] Ramadhanti, D., Larasati, A., Muid, A. and Mohamad, E., 2023, February. Building customer churn prediction models in Indonesian telecommunication company using decision tree algorithm. In AIP Conference Proceedings (Vol. 2654, No. 1). AIP Publishing.

[10] Shen, T.J. and Shibghatullah, A.S.B., 2022. Developing Machine Learning and Deep Learning Models for Customer Churn Prediction in Telecommunication Industry.

[11] Zhao, Y., Shao, Z., Zhao, W., Han, J., Zheng, Q. and Jing, R., 2023. Combining unsupervised and supervised classification for customer value discovery in the telecom industry: a deep learning approach. Computing, pp.1-23.

[12] Nalatissifa, H. and Pardede, H.F., 2021. Customer Decision Prediction Using Deep Neural Network on Telco Customer Churn Data. Jurnal Elektronika Dan Telekomunikasi, 21(2), pp.122-127.

[13] 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.

[14] Xu, J., Liu, J., Yao, T. and Li, Y., 2023. Prediction and Big Data Impact Analysis of Telecom Churn by Backpropagation Neural Network Algorithm from the Perspective of Business Model. Big Data.

[15] Chinnaraj, R., 2022. Bio-Inspired Approach to Extend Customer Churn Prediction for the Telecom Industry in Efficient Way.

[16] Pustokhina, I.V., Pustokhin, D.A., Nguyen, P.T., Elhoseny, M. and Shankar, K., 2021. Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector. Complex & Intelligent Systems, pp.1-13.

[17] Aghakhani Bezdi Langari, A. and Hasani, A., 2023. Customer Churn Analysis Based on The Datamining Approach: Hybrid Algorithm Incorporates Decision Tree and Bayesian Network. New Marketing Research Journal.

[18] Liu, Y., Shengdong, M., Jijian, G. and Nedjah, N., 2022. Intelligent prediction of customer churn with a fused attentional deep learning model. Mathematics, 10(24), p.4733.

[19] Sinsomboonthong, S., 2021. Efficiency comparison in prediction of normalization with data mining classification. diabetes, 768(537), p.231.

[20] Alharbi, A.H., Abdelhamid, A.A., Ibrahim, A., Towfek, S.K., Khodadadi, N., Abualigah, L., Khafaga, D.S. and Ahmed, A.E., 2023. Improved Dipper-Throated Optimization for Forecasting Metamaterial Design Bandwidth for Engineering Applications. Biomimetics, 8(2), p.241.

[21] Méndez, M., Merayo, M.G. and Núñez, M., 2023. Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model. Engineering Applications of Artificial Intelligence, 121, p.106041.

[22] Ahmed Ali Agoub, R., Hançerlioğullari, A., Rahebi, J. and Lopez-Guede, J.M., 2023. Battery Charge Control in Solar Photovoltaic Systems Based on Fuzzy Logic and Jellyfish Optimization Algorithm. Applied Sciences, 13(20), p.11409.

[23] Brandusoiu, G. Toderean, H. Beleiu, Methods for churn prediction in the pre-paid mobile telecommunications industry, in 2016 International Conference on Communications (COMM), (2016), 97–100. https://doi.org/10.1109/ICComm.2016.7528311.

[24] 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.140214
Received: July 19, 2023 Revised: November 29, 2023 Accepted: January 21, 2024