Adagrad Optimizer with Elephant Herding Optimization based Hyper Parameter Tuned Bidirectional LSTM for Customer Churn Prediction in IoT Enabled Cloud Environment

S. Venkatesh
Assistant Professor, Department of Computer Science, Government Arts and Science College, Perumbakkam, Chennai. E-mail: venkas76@gmail.com

Dr.M. Jeyakarthic
Assistant Director (Academic), Tamil Virtual Academy, Chennai.
E-mail: jeya_karthic@yahoo.com

Received July 19, 2020; Accepted September 24, 2020
ISSN: 1735-188X
DOI: 10.14704/WEB/V17I2/WEB17057

Abstract

At recent times, customer churn is an important activity in quickly developing industries such as telecom, banking, e-commerce, etc. Earlier studies revealed that the cost of getting a new customer is considerably higher than the cost of retaining the existing ones. Therefore, it becomes essential to predict the nature of customer churn for retaining the customers to a greater extent. The advent of deep learning (DL) models have begun to be employed for efficient CCP. This paper presents a new Adagrad Optimizer with Elephant Herding Optimization (EHO) based Hyper-parameter Tuned Bidirectional Long Short Term Memory (AG-EHO-BiLSTM) for CCP in Internet of Things (IoT) enabled Cloud Environment. The proposed AG-EHO-BiLSTM model initially acquires the customer data using its devices like smart phones, laptop, smart watch, etc. Next, the gathered data will be classified by the use of Bi-LSTM model, which determines the customers as churner or non-churner. The efficiency of the Bi-LSTM model can be increased through hyper parameter tuning techniques, namely Adagrad optimizer and EHO algorithm to optimally select the parameter values namely learning rate, number of hidden layer and epochs. The performance validation of the AG-EHO-BiLSTM model takes place on benchmark dataset and the simulation outcome reported the supremacy of the AG-EHO-BiLSTM model over the comparative methods.

Keywords

Customer Churn Prediction, Learning Rate Scheduler, Hyperparameter Tuning, EHO Algorithm.

Introduction

Software as a Service (SaaS) is an evolving method where users started subscribing the services rather than purchasing a license possessing software. The services are offered for a specific duration like monthly or yearly services that are similar to mobile phone subscription such as a post paid or prepaid services. In addition to that, SaaS method offered a huge set of advantages to the users. The lowest subscription cost provides several end products for the firms that could not brought by the subscriber. System administration was monitored by the seller that leads to cost limitation and training duration of the subscribers. The monthly revenue stream of services is based on the subscriptions that are stimulated by the sellers; but the dependability of the sellers for upgrading the services to satisfy the user by making a business in an exquisite manner. In case of buying a traditional software license, then the software is installed on the client's server and the seller is not aware of the client usage about the purchased software.

SaaS vendors utilized the applications deploying the software on servers are controlled by the sellers. It enables the seller to collect the valuable data that includes the data such as number of times the product is utilized, when, duration of usage, and who employs the product. The firms like telecom and finance, based on the pattern that distinguishes user churn and satisfaction. It has the capacity of determining the similar features of the clients. Customer churn has been employed widely in many firms. Therefore, SaaS is incorporated with the basic idea for the prevention of customer churns. However, SaaS industries allocated several commonalities with the telecom infrastructure. There are some main aspects which are the reason for the drastic SaaS market development and increase the accessibility and the expansion of Cloud Computing (CC).

Customer churn is explained as the problem to be observed in recent growing and challenging industry. The major goal of this industry was evolved as acquiring new clients rather retaining the pervious clients is extremely possible that requires less cost [1]. Maintaining the old clients leads to maximize the demand and restrain the cost of the retailing related to the obtaining new users. Specially, it is ended to customer churns and prediction activities must be a necessary part of several industries which supports in static decision making and designing procedure. Maintaining the user is the essential process of Customer Relationship Management (CRM). It directs to the evolution of various types of methodologies, that helps more essential function in the design of prediction and classification models. User churn is described as the change from one service provider to another opponent in the sector. It remains as a significant issue in the highly challenging marketplace, particularly supervised in telecom industry [2]. Usually, 3 types of Customer churns are present and they are defined here [3]. Active churner (Volunteer): User who desires to abscond the deal and moves to another vendor. Passive churner (Non-Volunteer): If an organization wants to end the subscription to a client. Rotational churner (Silent): Users who ended the dealing without any notice to both parties like user and the organization. The active and passive churners could be simply calculated utilizing the earlier techniques according to the Boolean class value, but final churn is complicated as these kinds of customers may come in future. The dealers must have the aim of minimizing the churn proposition since it is the prominent feature that the old customers are extremely convincing resources for the organization rather that gathering fresh customers.

Nowadays, industries are more focused on the permanent relation with the clients and observe the activities of the customers at all the time with the use of knowledge discovery in Database (KDD) technique [4-7] for the extraction of the concealed relationship of the individual and features in the huge amount of data. The aspects are impressed by several firms for CRM examination that have the tendency in retaining the data of the customer. User related technique is usual, particularly in telecom industry, which finds the customer's activities based upon the past records stored in CRM. The data saved in CRM methods are converted into valuable data for the increasing problems of customer churn which detects user's churn behavior earlier to losing the user that increase the transaction and the efficiency [7].

Though various models have been available in the present, there is still a need to determine the customer churn with maximum detection rate. Since DL models find beneficial in several domains, this paper made an attempt to develop a new Adagrad Optimizer with Elephant Herding Optimization based Hyper-parameter Tuned Bidirectional LSTM (AG-EHO-BiLSTM) for CCP in IoT enabled Cloud Environment. The proposed AG-EHO-BiLSTM model involves four main processes, namely data acquisition, parameter tuning, learning rate scheduling, and classification. Once the data is gathered, it will be classified by the use of the Bi - LSTM model, which determines the customers as churner or non-churner. The efficiency of the Bi-LSTM model can be increased through hyper parameter tuning techniques, namely Adagrad optimizer and EHO algorithm to optimally select the parameter values namely learning rate, number of hidden layer and epoch count. Firstly, the EHO algorithm is used to tune the two parameters of Bi-LSTM namely number of hidden layer and neurons. Followed by, Adagrad optimizer is applied as a learning rate scheduler to determine the appropriate value of the learning rate with an aim of achieving maximum classifier performance. The performance validation of the AG-EHO-BiLSTM model takes place on benchmark dataset and the results are examined under diverse aspects.

Related Works

In [8], a statistical related predictive technique for the CCP is presented in telecommunication sectors by data mining (DM) and ML methods, namely logistic regression (LR) and decision tree (DT). The presented CCP model includes three phases, planning Web interface, feature extraction, and prediction. Data obtained by the DT is used for the estimation of all attributes where LR utilized maximum possibility estimation for transforming related variables to logistical ones. Therefore, they discovered DM is the most secure method for CCP. In [9], support vector machine (SVM) technique involved four kernel processes for estimating the user churn on the remote telecommunication database. Gain evaluation is used for the estimation process. By monitoring the obtained outcomes, it is observed that the predictions (i.e., for churn and non-churn) given by 'Polynomial kernel' has obtained the best outcomes with better accuracy.

In [10], boosting method is used to improve the efficiency of CCP technique. Besides, other authors who used to boost to enhance the accuracy of the given base technique, the available technique has separated the training data into two sets based on the weight defined by boosting technique as well as build individual technique for every cluster. A relative investigation showed that the result of logistic regression (LR) on the entire training datasets understands that the boosting provided a good separation and defined churners with higher risks. They proposed that the results of efficiency can be maximized with the support of other classification methods. A multilayer perceptron (MLP), neural network (NN) technique with back- propagation learning technique is used in [11] for churn prediction in the telecommunication sector. It used two versions on MLP- related techniques such as variation in error model and tuning of weight in ANN. In [12], a new attribute set is projected and then utilized seven methods (decision tree (DT), linear classifier (LC), SVM, MLP, DMEL, logistic regression (LR) for CCP. The relative investigation is done among the new attributes and existing attribute sets in terms of performance. They recognized that the new set of features with seven methods provided better outcomes for CCP comparing to already available feature set in the telecommunication sectors. The introduced novel feature sets posses few restrains and the methods required to be highly focused in future for the betterment in outcomes.

In [13], distance factor is used for the design of CCP model. The database was divided into two parts, one with high certainty other with low certainty. They discovered a technique based on the distance factors and its outcomes in different areas for the estimation of unpredicted certainty decision. In [14], a novel CCP design for telecom industries in South Asia is planned related to fuzzy classifiers. The method utilized VQNN, DWANN and FuzzyNN for predicting the churner from the customer file accurately. Many classification methods, namely NN, C4.5, SVM, Adaboost, gradient boosting, random forest, and linear regression are not same as the fuzzy classifier model used for finding the high accurateness.

The Proposed AG-EHO-BiLSTM Model

The workflow of the AG-EHO-BiLSTM model is shown in Fig. 1. Initially, customer data gets collected by the use of IoT devices and is sent to the cloud database server (CDS). Then, preprocessing takes place to remove the unwanted data. Then, the BiLSTM based classification process gets executed to classify the customer into churner and non-churner, which involves Adagrad and EHO based optimization processes. Besides, an alarm will be raised on the detection of churner customer. These processes are discussed in the subsequent sections.

1). Preprocessing

The preprocessing is carried out in 2 phases: noise elimination as well as format conversion. Initially, the original data are computed for removing missing values. Followed by, the categorical data would be transformed into mathematical ones to improvise the optimal prediction task. The categorical measure of state attributes are converted into mathematical data as well as binary values are converted as 0 and 1 correspondingly. After the completion of pre-processing, classification has been carried out using BiLSTM model.

2). Bi-LSTM Model

An LSTM depends upon the traditional RNN structure. It applies diverse models for computing hidden state which resolves the problem of Recurrent ANN of unable to manage long distance dependence. However, the best potential of LSTM is not learned by these models, even though the inherent merits of newly developed model. This LSTM approach is composed of a set of memory units with 3 gates and different performance. Under the application of text feature vector S as input and t^{th} word as a sample, and values of specific conditions of LSTM unit of t^{th} words are provided in the following: A specific processing expression is provided below, where σ is the sigmoid function, $\odot$ defines a dot multiplication.

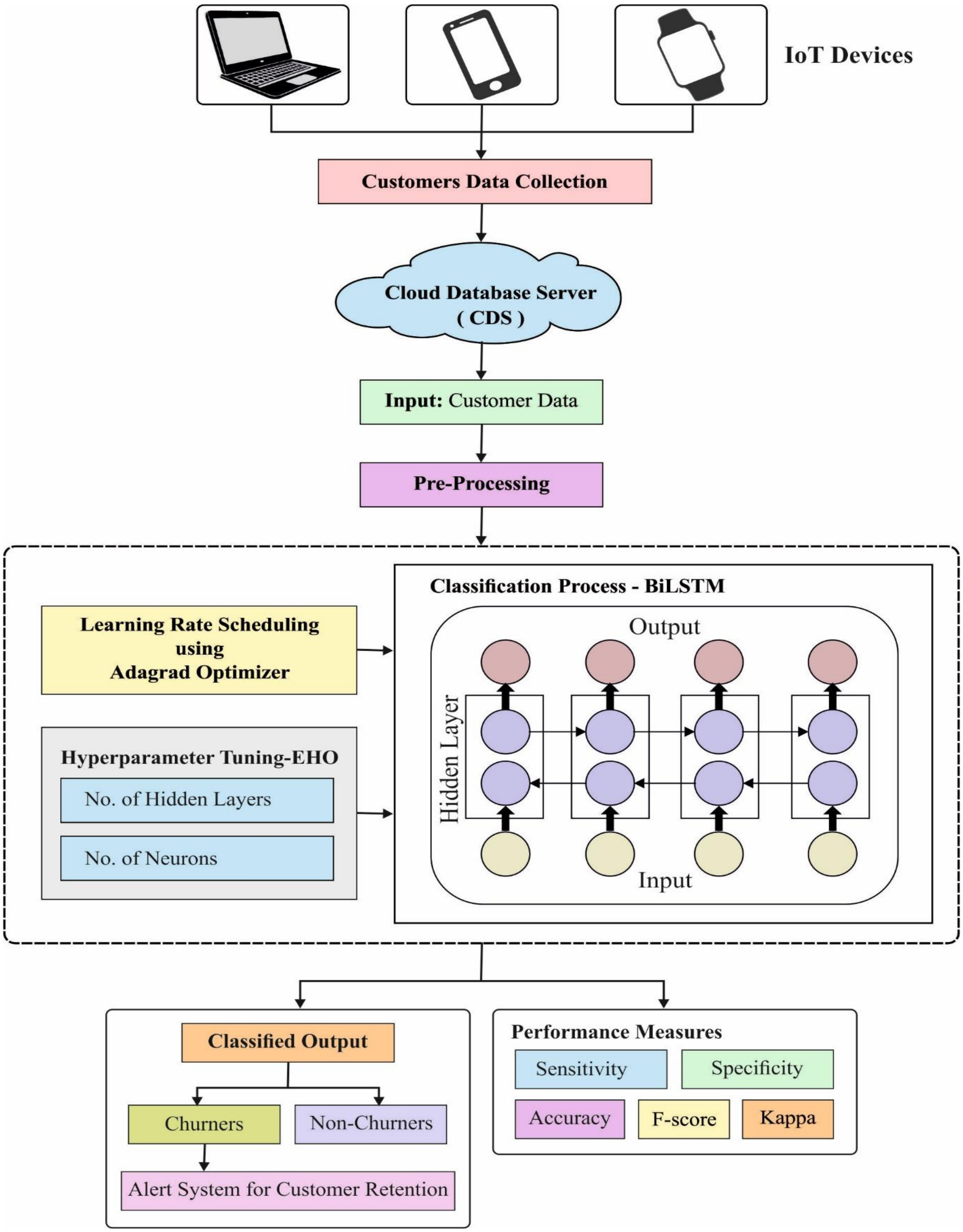

Fig. 1 Workflow of AG-EHO-BiLSTM model

The f_t implies a forget gate:

$$f_t = \sigma\left(W_f w_t + U_f h_{t-1} + b_f\right) \ (1)$$

The i_t signifies an input gate:

$$i_t = \sigma(W_i w_i + U_i h_{t-1} + b_i) \ (2)$$

The $\tilde{c}_t$ Consumes the candidate memory cell condition recently, in which tanh defines the tangent hyperbolic function;

$$\tilde{c}_t = \tanh(W_c w_t + U_i h_{t-1} + b_c) \ \ (3)$$

The c_t denotes the state value of recent time in memory cell: the measure of f_t and i_t series from 0 to 1. The processing of $i_t \odot c_t$ implies that new data is stored in c_t from candidate unit $\tilde{c}_t$. The function of $f_t \odot c_{t-1}$ indicates that the data is conserved and it is unwanted for memories c_{t-1}.

$$c_t = i_t \odot \tilde{c}_t + f_t \odot c_{t-1} \ (4)$$

The o_t indicates output gate:

$$o_t = \sigma(W_o w_t + U_o h_{t-1} + b_o) \ (5)$$

h_t implies hidden layer state at time t:

$$h_t = o_t \odot \tanh(c_t) \ \ (6)$$

The LSTM contains historical series of data and it is insufficient. If it applies upcoming data, then it is applicable to a series of roles. The bidirectional LSTM is comprised of a forward and backward LSTM layer [15]. The r function is provided in the following: the forward layer applies historical data of a series while backward layer consumes the upcoming data of the series. Followed by, these layers are joined to the same resultant layer. The major advantage of this model is that the series context data is regarded completely. It is considered that the input of time t is a word embedding w_t, at time $t-1$, the final outcome of forward hidden unit is $\vec{h}_{t-1}$, and simulation outcome of the backward hidden unit is $\overleftarrow{h}_{t+1}$, Finally, the result of backward as well as hidden unit at time t is same as given in the following:

$$\vec{h}_t = L\left(w_t, \vec{h}_{t-1}, c_{t-1}\right) \text{ (7)}$$

$$\overleftarrow{h}_t = L\left(w_t, \overleftarrow{h}_{t+1}, c_{t+1}\right) \text{ (8)}$$

Where $L\,(.)$ refers the hidden layer of the LSTM hidden layer A forward resultant vector is $\vec{h}_t \in R^{1\times H}$ as well as backward output vector is $\overleftarrow{h}_t \in R^{1\times H}$, which has to be combined to accomplish text feature. It is described that H shows the amount of hidden layer cells:

$$H_t = \vec{h}_t || \overleftarrow{h}_t \text{ (9)}$$

3). Adagrad based Learning Rate Scheduler

When the DL method is trained, it is used for limiting the learning rate (γ_t) while there is a rapid development in the training phase. The weight count is improved while training a step size or "learning rate." Specifically, learning rate is a modeling hyper-parameter used for the training NN with low positive value, from 0.0 and 1.0. The learning rate balances the path for resolving these issues. Minimum learning rates require higher training epochs which offer small alterations for the weights. The learning rates are intended to offer massive changes and require low training epochs. The task of simulating learning rate is highly complex. A maximum learning rate results in divergent training process while minimum learning rate tends to slow convergence. An effective result can be attained by triggering various learning rate while training.

Adagrad is a model used for gradient-relied optimization and it applies learning rate for parameters. It performs tiny updates where the parameters related with prominent features as well as larger updates for parameters are correlated with dissimilar features. For this sake, it is highly applicable to deal with sparse data. Adagrad enhances the efficiency of stochastic gradient descent (SGD) and applied for training large-scale neural nets. Adagrad optimizer is gradient relied optimization models which operate quite well for sparse gradients. The learning rate mimics how much the parameter enables to apply opposite direction of a gradient estimate (g). It applies the learning rate, according to the attributes.

$$\Theta_{t+1} = \Theta_t - \frac{\alpha}{\sqrt{\varepsilon + \sum g_t^2}} \odot g_t \ (10)$$

The fundamental function for parameter update is depicted in Eq. (10)) where θ_t shows parameter at time t, α defines the learning rate, g_t signifies gradient estimate, and $\odot$ refers element wise multiplication.

4). EHO based Hyperparameter Tuning

Hyper-parameter is a parameter which has to be selected manually earlier for training. The term 'hyper-'is to classify hyper-parameter to parameter which has been modified independently by using optimization algorithms at the time of the training phase. But, when the hyper-parameter may be complex for a set as it is too small, the parameter update would be slow and it consumes maximum duration to accomplish reasonable loss. Else, when it is fixed with large value, then the parameter is moved against a function and no data loss exists. The high-dimensional non-convex behavior of NN optimization results in various sensitivity of all dimensions. The major aspect is to normalize the hyper-parameters of Bi-LSTM classification with the application of EHO algorithm and attains best function on CCP. Here, the parameters are, batch size as well as the number of hidden neurons. This EHO method is initialized from primary solutions that has been produced randomly and tries to maximize the accuracy of CCP method iteratively till reaching the termination condition. Hence, the fitness function is Bi-LSTM networks that are in-charge to perform the estimation and provide CCP accuracy.

The EHO model is defined by the application of giving simplified rules [16] and the working process is shown in Fig. 2:

1. Elephants from diverse clans reside together and ruled by a leader named as matriarch. A clan has a limited number of elephants. In case of modeling, consider that every clan is composed of the same count of elephants.
2. The locations of elephants in a clan are often changed according to the matriarch. EHO models this nature using an updating operator.
3. Adult male elephants leave the groups and resides lonely. Also, the fixed numbers of male elephants leave the clans. Finally, EHO labels the updating task with the help of separating operator.
4. Basically, the matriarchs in all clans are eldest female elephant. In order to resolve the optimization issues, the matriarch is assumed as fittest elephant in a clan.

a). Clan Updating Operator

Let an elephant clan is c_i. Followed by, upcoming location of an elephant is j in the clan, which has been upgraded by Eq. (11).

$$x_{new,ci,j} = x_{ci,j} + \alpha \times (x_{best,ci} - x_{ci,j}) \times r, (11)$$

where $x_{new,ci,j}$ depicts the upgraded position, and $x_{ci,j}$ implies advanced location of an elephant j in clan ci. $x_{best,ci}$ represents a matriarch of the clan c_i; where female one is best elephant in the clan. The scale factor $\alpha \in [0,1]$calculates the power of a leader of c_i on $x_{ci,j}$. $r \in [0, 1]$, where it is a stochastic distribution that offers an enhancing objective for population diversity in secondary searching stage. Recently, a uniform distribution has been employed.

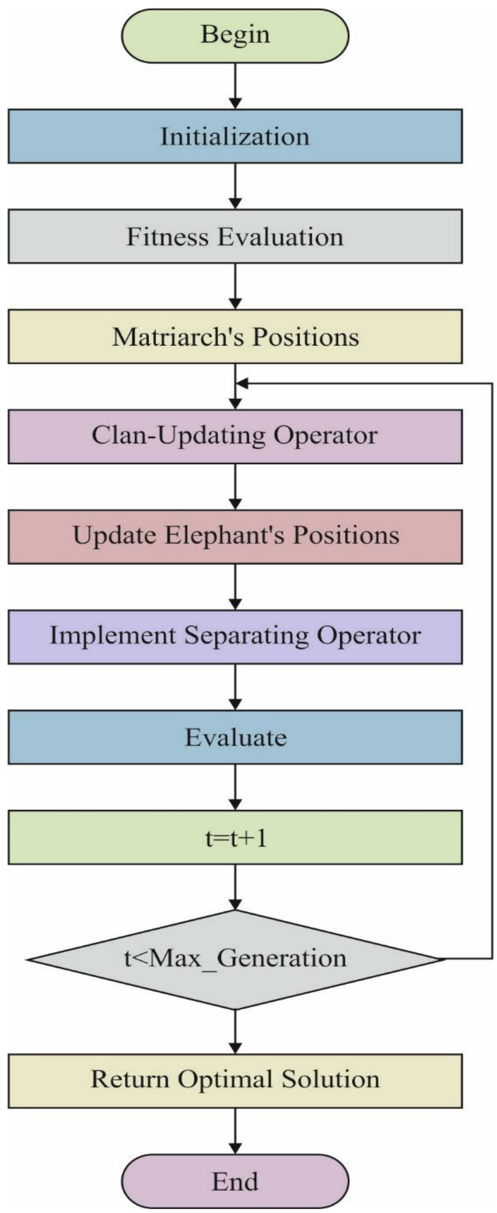

Fig. 2 Flowchart of EHO algorithm

It has to be pointed that $x_{ci,j} = x_{best,ci}$, refers that the matriarch in the clan could not be upgraded using (11). This situation can be avoided by upgrading the best elephant with the help of the given function:

$$x_{new,ci,j} = \beta \times x_{center,ci}, (12)$$

where the power of $x_{center,ci}$ on $x_{new,ci}$, is regularized by $\beta \in [0,1]$.

The information from every individual in clan c_i has been applied for developing a novel individual $x_{new,ci,j}$. The middle of clan ci, $x_{center,ci}$, could be evaluated for d-th dimension by D calculations, where Ddenotes the overall dimension, as given in the following:

$$x_{center,ci,d} = \frac{1}{n_{ci}} \times \sum_{j=1}^{n_{ci}} x_{ci,j,d} \quad (13)$$

In this approach, $1 \leq d \leq D$implies the d-th dimension, n_{ci}signifies count of individuals in ci, and $x_{ci,j,d}$ indicates dth dimension of individual $x_{ci,j}$.

Algorithm 1 offers the pseudocode for updating operator.

Algorithm 1: Clan updating operator

```
Begin
    for ci = 1 to nClan (for every clans in elephant
        population) do
     for j = 1 to n_ci (for all elephant individuals in clanci)
        do
            Update x_ci,j and generate x_new,ci,j according to (11).
            if x_ci,j = x_best,ci then
                Update x_ci,j and produce x_new,ci,j based on the Error! Reference
            source not found.
                end if
        end for j
    end for ci
End
```

b). Separating Operator

In every clan, male elephants leave their family group and live lonely once it is matured. The process of separation is labeled as a separating operator while the optimization issues are resolved. To enhance the searching potential of EHO model, consider that an elephant with the inferior fitness would execute a separating operator for all generations as illustrated in Eq. (14).

$$x_{worst,ci} = x_{min} + (x_{max} - x_{min} + 1) \times rand \quad (14)$$

where x_{max} and x_{min} are upper and lower bound, correspondingly, of location of an elephant. $x_{worst,ci}$ defined as a poor individual elephant in clan ci. $rand \in [0,1]$ implies a type of stochastic distribution, as well as uniform distribution within [0,1] as applied recently.

Then, the separating operator has been developed, as depicted in Algorithm 2.

Algorithm 2: Separating operator
Begin
for $ci = 1$ to $nClan$(all of the clans in the elephant population) do
Replace the inferior elephant individual in clan ci using (14).
end for $\boldsymbol{ci}$
End

c). Schematic Presentation of the Basic EHO Algorithm

In case of EHO, similar to alternate meta-heuristic models, a kind of elitism principle has been applied for securing the fittest elephant individuals by the prevention of clan updating and separating operators. Initially, the better elephant individuals are secured, and the poor ones are interchanged by secured best elephant individuals finally. The elitism assures that the later elephant population is not worse than previous one. The schematic representation is consolidated as provided in Algorithm 3.

Algorithm 3: Elephant Herd Optimization (EHO)
Begin Step 1: Initialization. Fix the generation counter $t = 1$. Upload the population P of NPelephant individuals arbitrarily, with uniform distribution in a search space. Fix the count of kept elephants $nKEL$, the higher generation $MaxGen$, the scale factor α and β,the number of clan $nClan$,and the count of elephants for the ci-th clan n_{ci}. Step 2: Fitness estimation. Determine each elephant individual on the basis of its position. Step 3: While $t < MaxGen$do the following: Arrange all elephant individuals based on their fitness. Secure the $nKEL$elephant individuals. Execute the clan updating operator as illustrated in Algorithm 1. Implement the separating operator as presented in Algorithm 2. Determine the population on the basis of newly upgraded locations. Swap the poor elephant with the $nKEL$saved ones. Upgrade the generation counter, $t = t + 1$. Step 4: End while Step 5: Output the best solution. End.

Experimental Validation

For examining the efficiency of the proposed CCP technique, an open result investigations are carried out on a standard dataset and the outcomes are evaluated by various measures like accuracy, specificity, kappa, F-score and sensitivity. For comparison purposes, a set of methods, namely Vote, SVM, Naïve Bayes (NB), DT, Neural Network (NN), LogitBoost, LR, optimal genetic algorithm with SVM (OGA-SVM), and Preprocessing with Adaptive Gain with Back Propagation Neural Networks (P-AGBPNN) [17].

1). Dataset used

For the experiments, a standard telecom data is utilized. It includes a sum of 3333 samples with the presence of 21 attributes. It comprises 21 attributes and 2 class labels. The 14.49% sample set comes under the positive class and the remaining of 85.51% of samples fall into negative class. The dataset particulars are given in Table 1. On the other hand, the feature list occurs in the dataset is displayed and the frequency allocation of samples in the real dataset is portrayed in Fig. 3.

Table 1 DataSet Description

Description	Dataset
No. of Instances	3333
No. of Features	21
No. of Class	2
Percentage of Positive Samples	14.49%
Percentage of Negative Samples	85.51%
Data sources	[18]

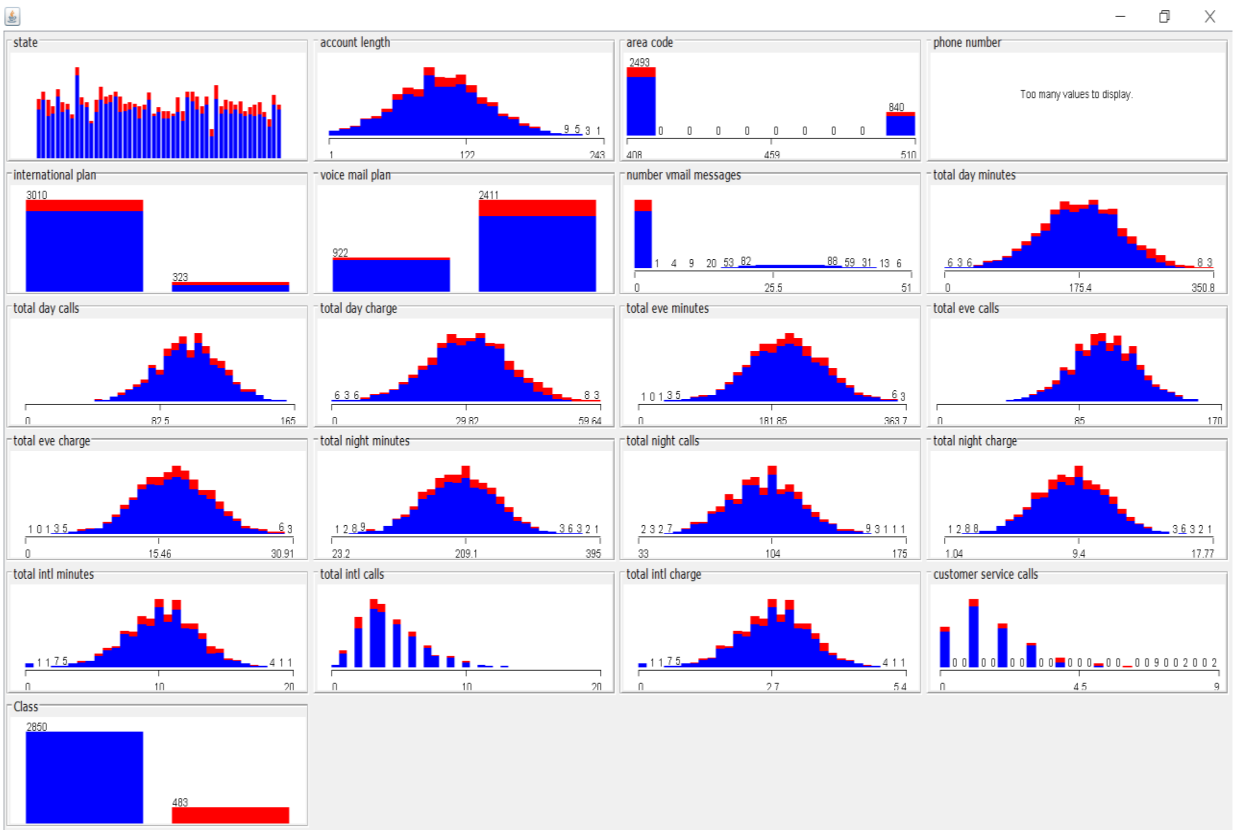

Fig. 3 Frequency Distribution of Dataset for All Attributes

2). Results Analysis

Table 2 tabulates the confusion matrix obtained through BiLSTM and technique and AG-EHO-BiLSTM model. The values in the table showed which the BiLSTM technique has properly calculated 2791 samples as non-churn and 416 samples as churn. Similarly, the combination of Adagrad and EHO techniques in AG-EHO-BiLSTM model has exhibited high results by efficiently predicting 2827 samples as non-churn and 451 samples as churn. These values established that the efficiency of the classifiers starts improving through the involvement of parameter tuning processes.

Table 2 Confusion Matrix of Proposed Methods

Methods	Descriptions	Values
BiLSTM	Number of Instances Predict Non-Churners as Non-Churners (TP)	2791
	Number of Instances Predict Churners as Churners (TN)	416
	Number of Instances Predict Non-Churners as Churners (FP)	59
	Number of Instances Predict Churners as Non-Churners (FN)	67
AG-EHO-BiLSTM	Number of Instances Predict Non-Churners as Non-Churners (TP)	2827
	Number of Instances Predict Churners as Churners (TN)	451
	Number of Instances Predict Non-Churners as Churners (FP)	23
	Number of Instances Predict Churners as Non-Churners (FN)	32

Table 3 and Fig. 4 give a clear investigation of the prediction outcome provided by the proposed with the compared techniques. The figure indicated the sensitivity and the specificity investigation of the existing technique. The experimental values exposed that the DT technique offered bad efficiency with the minimum 76.47% sensitivity value and 79.49% specificity value. At the same time, the NN technique showed excellent performance than the DT technique of obtaining somewhat better sensitivity of 81.75% and specificity of 94.7%. Similarly, the Vote has shown low sensitivity of 85.51%. The Logit boost and LR technique have exposed closer sensitivity values of 86.31% and 86.52%, and specificity of 43.82% and 45.71% correspondingly. Also, the NB technique has attained good efficiency with the earlier technique by providing maximum sensitivity of 92.21% and specificity of 57.98% correspondingly. The OGA-SVM and P-AGBPNN exhibited nearer sensitivity values of 94.5% and 95.5%, and specificity values of 66.06% and 70.49% correspondingly. Although the BFS-AFSA-MLP technique (i.e. AFSA-MLP prior to FS) has confirmed effective outcome with the existing techniques with the sensitivity of 96.37% and specificity of 77.08%, the AFS-AFSA-MLP (i.e. AFSA-MLP following FS) has exhibited best performance that the earlier techniques with the high sensitivity of 97.28% and specificity of 81.53%. At the end, the near optimal classifier results with the higher sensitivity of 97.66% and specificity of 87.58% has been achieved by BiLSTM whereas the presented AG-EHO-BiLSTM model has shown outstanding results with the highest sensitivity and specificity of 98.88% and 95.15% respectively.

Table 3 Performance Evaluation of Existing with Proposed AG-EHO-BiLSTM Method

Methods	Sensitivity	Specificity	Accuracy	F-Score	Kappa
AG-EHO-BiLSTM	98.88	95.15	98.35	99.04	93.29
BiLSTM	97.66	87.58	96.22	97.79	84.64
AFS-AFSA-MLP	97.28	81.53	94.93	97.03	79.80
BFS-AFSA-MLP	96.37	77.08	93.52	96.20	74.07
P-AGBPNN	95.50	70.49	91.71	95.13	67.20
OGA-SVM	94.50	66.06	90.27	94.30	61.17
LR	86.52	45.71	85.34	91.90	11.76
LogitBoost	86.31	43.82	85.18	91.89	9.56
NN	81.75	94.70	93.20	73.20	-
DT	76.47	79.49	77.90	78.50	-
Naive Bayes	92.21	57.98	87.64	92.81	48.44
SVM	93.46	60.13	88.30	93.19	61.40
Vote	85.51	-	85.51	92.18	0

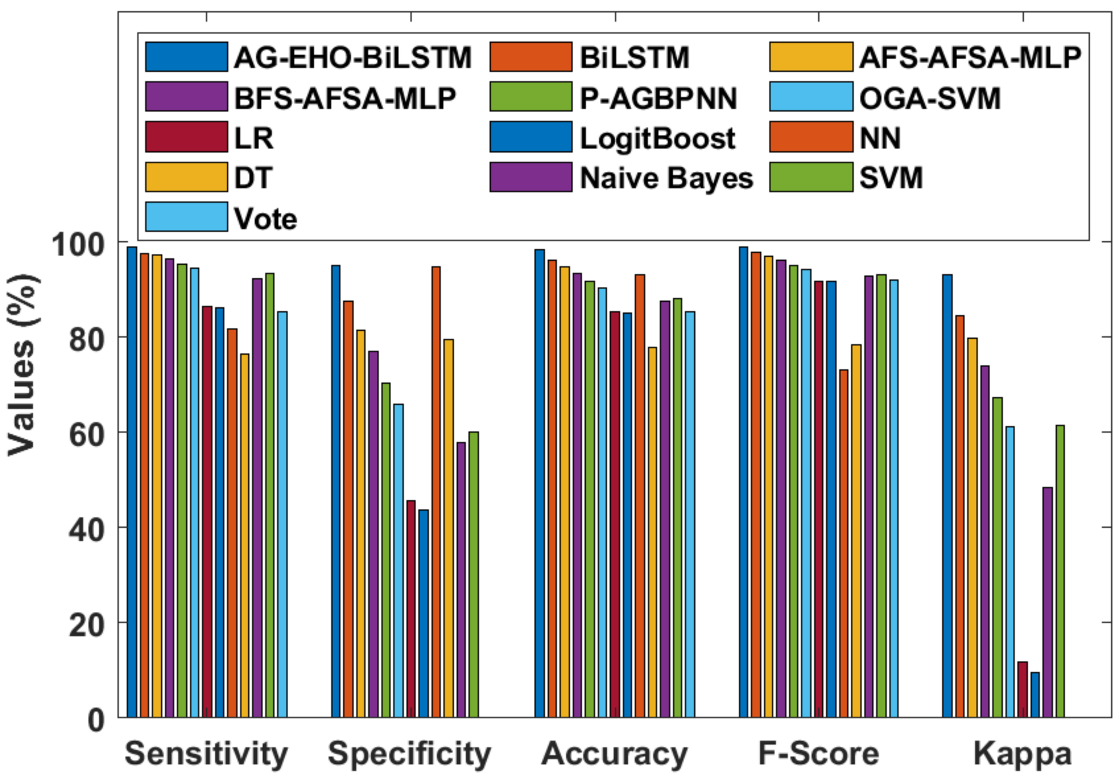

Fig. 4 Comparative analysis of AG-EHO-BiLSTM model with existing methods

The figure also demonstrated that the F-score and kappa investigation of the proposed technique. The experimental values showed that the NN technique provides very bad efficiency with the least 73.20% of F-score. In the meantime, the DT technique has shown better efficiency than NN technique by achieving an effective F-score of 78.50%. Both the Logit boost and LR approaches have exhibited nearer F-score values of 91.89% and 91.90%, and kappa values of 9.56% and 11.76% respectively. At the same time, the vote exhibited low F-score of 92.18% and kappa value of 0%. Simultaneously, the NB technique has obtained reasonable efficiency with the previous techniques by offering high F-score of 92.81% and kappa of 48.44% correspondingly. Similarly, the SVM technique has shown gradually lower F-score of 93.19% and kappa of 61.40% correspondingly. Likewise, the OGA-SVM and P-AGBPNN have exhibited closer F-score values of 94.30% and 95.13%, and kappa values of 61.17% and 67.20% correspondingly. Though the BFS-AFSA-MLP technique (i.e. AFSA-MLP prior to FS) has depicted moderate outcomes compared to previous models with the F-score of 96.20% and kappa of 74.07%, the improved AFS-AFSA-MLP (i.e. AFSA-MLP following FS) has outperformed the compared techniques with the high F-score of 97.03% and kappa of 79.80%. Although the BiLSTM model has exhibited competitive performance with the F-score of 97.79% and kappa 84.64%, the presented AG-EHO-BiLSTM Method has showcased better performance with the maximum F-score of 99.04% and kappa 93.29%.

The figure displayed the predictive accuracy of the proposed and existing techniques. The obtained values exhibited that the DT technique provides poor outcome with the least accuracy value of 77.90%. Simultaneously, the Logit boost technique performed well than the DT technique by attaining the normal accuracy value of 85.18%. The Vote and LR techniques have applied nearer similar accuracy value of 85.51% and 85.34% respectively. At the same time, the NB has depicted minimum accuracy value of 87.64%. Likewise, the SVM technique has attained considerable efficiency with the previous techniques by providing high accuracy value of 88.30%. In line with this, the OGA-SVM and P-AGBPNN models have portrayed nearer accuracy values of 90.27% and 91.71% respectively. Similarly, the NN technique has showcased the effective accuracy value of 93.20%. Though the BFS-AFSA-MLP, AFS-AFSA-MLP and BiLSTM techniques demonstrated comparatively convenient outcome with the earlier technique with the accuracy of 93.52%, 94.93% and 96.22%, the proposed AG-EHO-BiLSTM Method has shown superior results to previous techniques with the maximum accuracy value of 98.35%.

Table 4 implies the comparison analysis with current models for given dataset by means of Accuracy as well as the F-Score. Fig. 5 depicts a brief comparative analysis of the recent models for applied dataset in terms of accuracy. The experimental measures showed that the DT approach provides ineffective function with least accuracy value of

77.90%. Simultaneously, the LDT/UDT-1 and LDT/UDT-2 frameworks have outperformed than DT method of achieving better and closer accuracy measures of 84%. Meantime, the LDT/UDT-10 has exhibited minimal accuracy value of 84.30%. The LDT/UDT-8 and LDT/UDT-6 approaches have showed closer and similar accuracy values of 84.63% and 84.67% respectively. In line with this, the LDT/UDT-4 and LDT/UDT-9 models have attained considerable performance than the previous schemes by reaching higher accuracy values of 84.75% and 84.78% correspondingly. Along with that, the LDT/UDT-7 and LogitBoost methodologies have demonstrated maximum accuracy values of 84.86% and 85.18% respectively. On the same way, the LDT/UDT-3, LR and LDT/UDT-5 methods showcased closer accuracy values of 85.33%, 85.34% and 85.40% respectively. Simultaneously, the OGA-SVM, P-AGBPNN and NN methods have exhibited best performance result with accuracy values of 90.27%, 91.71% and 93.20% correspondingly. Even though the BFS-AFSA-MLP, AFS-AFSA-MLP and BiLSTM methodologies have depicted maximum outcome than the earlier models with the accuracy values of 93.52%, 94.93 and 96.22%, the projected AG-EHO-BiLSTM model has performed quite-well when compared to the existing approaches with greater accuracy of 98.35%.

Table 4 Comparison with Recent Methods for Applied Dataset in terms of Accuracy and F-Score

Methods	Accuracy	F-Measure
AG-EHO-BiLSTM	98.35	99.04
BiLSTM	96.22	97.79
AFS-AFSA-MLP	94.93	97.03
BFS-AFSA-MLP	93.52	96.20
P-AGBPNN	91.71	95.13
OGA-SVM	90.27	94.30
LR	85.34	91.90
LogitBoost	85.18	91.89
NN	93.20	73.20
DT	77.90	78.50
LDT/UDT-1	84.00	57.89
LDT/UDT-2	84.00	54.29
LDT/UDT-3	85.33	54.17
LDT/UDT-4	84.75	55.47
LDT/UDT-5	85.40	56.29
LDT/UDT-6	84.67	54.90
LDT/UDT-7	84.86	57.60
LDT/UDT-8	84.63	58.02
LDT/UDT-9	84.78	56.23
LDT/UDT-10	84.30	56.02

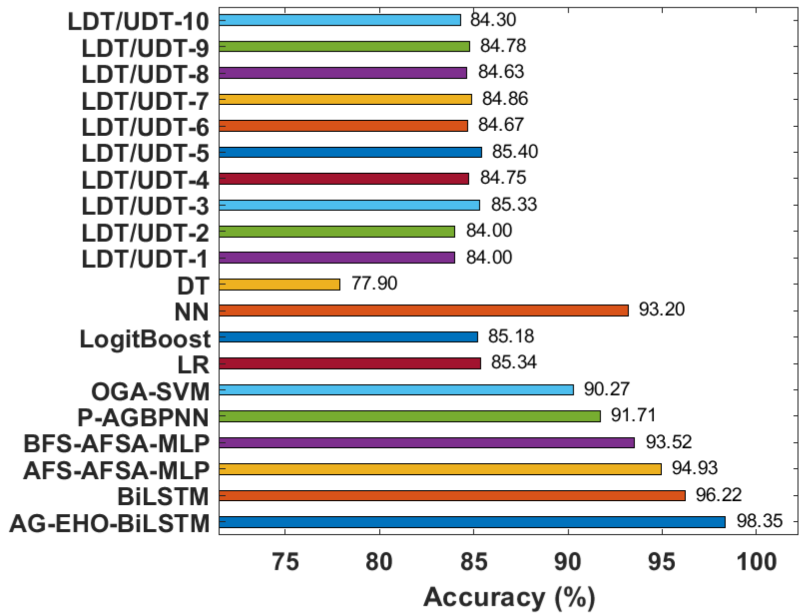

Fig. 5 Comparison with recent methods in terms of Accuracy

Fig. 6 gives a detailed investigation compared with the existing technique for provided dataset in accordance with the F-measures. The experimental values explained that the LDT/UDT-3 technique gives bad efficiency with the low F-measures value of 54.17%. In the meantime, the LDT/UDT-2 and LDT/UDT-6 techniques showed higher efficiency than the LDT/UDT-3 technique by achieving medium F-measure values of 54.29% and 54.90% respectively. Simultaneously, the LDT/UDT-4 has illustrated low F-measure value of 55.47%. The LDT/UDT-10 and LDT/UDT-9 models have depicted closer identical F-measure values of 56.02% and 56.23% correspondingly.

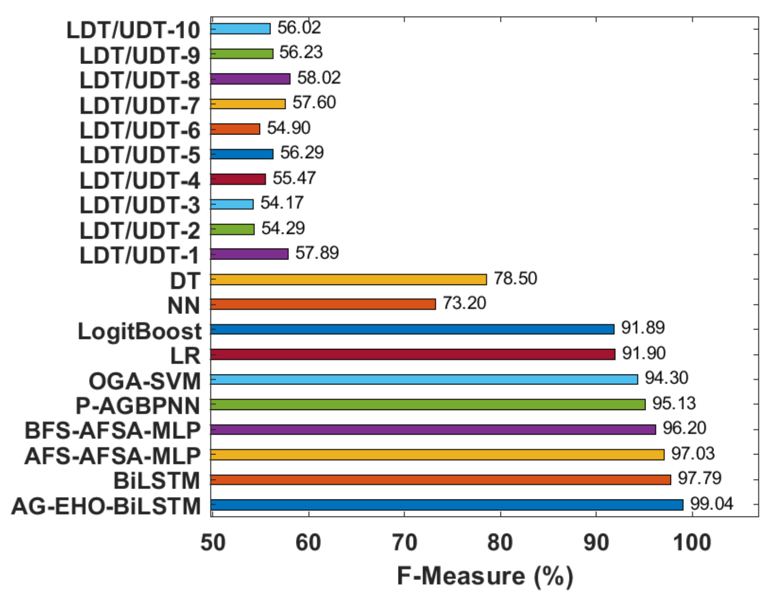

Fig. 6 Comparison with recent methods in terms of F-Measure

In line with this, the LDT/UDT-5 and LDT/UDT-7 techniques have obtained realistic efficiency with the previous techniques by providing high F-measure values of 56.29% and 57.60% respectively. Likewise, the LDT/UDT-1 and LDT/UDT-8 approaches have displayed effective F-measure values of 57.89% and 58.02% correspondingly. At the same time, the NN, DT and Logit boost have showed near F-measure values of 73.2%, 78.5% and 91.89% correspondingly. Similarly, the LR, OGA-SVM and P-AGBPNN technique has demonstrated effective outcomes along with F-measure values of 91.90%, 94.3% and 95.13% correspondingly. Though the BFS-AFSA-MLP, AFS-AFSA-MLP and BiLSTM techniques have exhibited better efficiency when related to conventional techniques with the F-measure value of 96.20%, 97.03% and 97.79% respectively. However, the presented AG-EHO-BiLSTM model has demonstrated superior results with the maximum F-measure value of 99.04%.

Conclusion

This paper has presented a new AG-EHO-BiLSTM for CCP in IoT enabled Cloud Environment. The proposed AG-EHO-BiLSTM model initially acquires the customer data using IoT devices like smart phones, laptop, smart watch, etc. Next, the gathered data will be classified by the use of Bi-LSTM model, which determines the customers as churner or non-churner. The efficiency of the Bi-LSTM model can be increased through hyper parameter tuning techniques, namely Adagrad optimizer and EHO algorithm to optimally select the parameter values namely learning rate, number of hidden layer and epochs. The performance validation of the AG-EHO-BiLSTM model takes place on benchmark dataset and the simulation outcome reported the supremacy of the AG-EHO-BiLSTM model over the comparative methods.

References

Veeramanikandan, V., & Jeyakarthic, M. (2019). An Ensemble Model of Outlier Detection with Random Tree Data Classification for Financial Credit Scoring Prediction System. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(3).

Huang, B., Kechadi, M.T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), 1414–1425.

Lazarov, V., & Capota, M. (2007). Churn prediction. *Bus. Anal. Course. TUM Comput. Science*, *33*, 34.

Sharma, A., & Kumar, P. (2011). A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. *International Janitorial Cleaning Services Association Application*, *27*(11), 26-31.

Thirumalairaj, A., & Jeyakarthic, M. (2020). An Intelligent Feature Selection with Optimal Neural Network Based Network Intrusion Detection System for Cloud Environment. *International Journal of Engineering and Advanced Technology (IJEAT)*, *9*(3).

Kirui, C., Hong, L., Cheruiyot, W., & Kirui, H. (2013). Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. *International Journal of Computer Science Issues (IJCSI)*, *10*(2 Part 1), 165-172.

Lina, C.S., Gwo-Hshiung, T., & Yang Chieh, C. (2011). Combined rough set theory and flow network graph to predict customer churn in credit card accounts. *Expert System with Application*, *38*(1), 8–15.

NDMA. (2017). National Disaster Management Authority Annual Report.

Rogers, E.M. (2003). *Diffusion of innovations.* The Free Press, New York.

Fishbein, M., & Ajzen, I. (1977). *Belief, attitude, intention, and behavior: an introduction to theory and research.* Addison-Wesley, Reading.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, *13*, 319–340.

Venkatesh, V., Morris, M.G., Davis, G.B., & Davis, F.D. (2003). User acceptance of information technology: toward a unified view. *MIS quarterly*, *27*, 425–478.

Goodhue, D.L., & Thompson, R.L. (1995). Task-technology fit and individual performance. *MIS quarterly*, *19*, 213–236.

Delone, W.H., & McLean, E.R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, *19*(4), 9-30.

Long, F., Zhou, K., & Ou, W. (2019). Sentiment analysis of text based on bidirectional LSTM with multi-head attention. *IEEE Access*, *7*, 141960-141969.

Wang, G.G., Deb, S., & Coelho, L.D.S. (2015). Elephant herding optimization. *In IEEE 3rd International Symposium on Computational and Business Intelligence (ISCBI)*, 1-5.

Jeyakarthic, M., & 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), 386-399.

http://www.sgi.com/tech/mlc/db/