

Tomtit Flock Optimization with Deep Variational Autoencoder for Customer Churn Prediction in Telecommunication Industry

Vani Haridasan*, Kavitha Muthukumaran, Hariharanath K

SSN School of Management, Kalavakkam, Chennai, India. *Corresponding Author's Email: vaniharidasanssn2022@gmail.com

Abstract

The telecommunication sector becomes popular in recent years and service companies suffer from a loss of valued customers to competitors. In the operational area, customer churn is a challenging issue, which makes companies, loses customers, and therefore is of bigger issue for the industry. Recent technologies and competitors were developing quickly and churn prediction becomes essential for telecom companies. Customer Churn Prediction (CCP) models offer precise detection of significant churners and therefore a retaining solution can be offered to them. Data analytics and machine learning (ML) models can be designed and used for churn investigation to reduce customer churns and improve profit. In this aspect, this article investigates a novel Tomtit Flock Optimization with Deep Variational Autoencoder for Customer Churn Prediction (TFODVAE-CCP) in the telecommunication sector. The presented TFODVAE-CCP technique mainly concentrates on the identification and classification of customer churns. To accomplish this, the presented TFODVAE-CCP technique performs data pre-processing at the primary stage. Next, the TFODVAE-CCP technique employs TFO algorithm to elect an optimal subset of features and thereby improve classification accuracy. Finally, the DVAE method can be leveraged for classifying churning/non-churning. A detailed set of experimental analyses is made on telecommunication churn dataset and the results illustrate the betterment of the TFODVAE-CCP method over other existing models.

Keywords: Customer Churn Prediction (CCP), Intelligent Models, Machine Learning (ML), Metaheuristics, Feature Selection, Telecommunication.

Introduction

Telecommunications is one of the rapidly growing industries in several nations. The average annual churn percentage of the telecommunication sector lies in-between 20 to 40%, which results in massive loss of revenue (1). An individual has various choices, and they will select the firms which would present them with less expensive and better-quality services (2). To withstand this highly competitive industry, telecommunication companies are in a need to formulate techniques to allure new customers or raise customer retention rates. It should be pointed out that acquisition of novel users would cost comparatively more than retention of an existing one. Thus, customer churn prediction (CCP) is becoming a famous research region as it has the potential to aid telecommunication companies in identifying individuals with higher potentiality of dismissing their contracts (3). After that, the companies analyze the situations and come up with appropriate solutions for these kinds of

users to employ to hold them. It was of critical importance for the development of the businesses over the longer term. Churn defines the consumer who will be transferring from single supplier of telephone services to any other (4). The telecommunication sector encounters main difficulty associated with user turnover; user does not extend their existing relationship with the network or business. This issue does not just affect the rapid evolution of a company; however, it also affects incomes (5). Thus, several CCP methods were enforced but do not provide favorable CCP act; this is because of potential variant parameters affecting customer churn (CC), which is unexplored till now. Customers can be churned for several reasons: the firm does not connect with their clients, grievance is not handled on time, adverse comments on social networking sites, press, and so on, unfulfilled with the services presented, and present software do not satisfy its requirements, the competitors give

This is an Open Access article distributed under the terms of the Creative Commons Attribution CC BY license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

(Received 13th March 2024; Accepted 23rd July 2024; Published 30th July 2024)

the novel product a superior price and quality, or not support advanced network kinds (6G, 4G, 5G). In many instances, the company is not aware of the reasons why the client is cancelling their service. In such a situation, it's possible that studying the information may be beneficial in comprehending and anticipating the CC so that you can avoid it happening (6). Managing the churning prediction issue can be accomplished by the application of a variety of Machine Learning (ML) strategies (7). Current research demonstrates that a viable solution for churn prediction must be able to efficiently harness a vast volume of past data in order to successfully execute churner detection. Yet, because there were additional restrictions placed on the prevalent methods, it is not possible to carry out churn estimation in a manner that is both proficient and precise to the greatest extent (8). A huge size of data was produced in telecommunication sector which has values which are missing. Estimation of such type of data leads to erroneous or poor results for predictive techniques in works. Data preprocessing was now carried out for addressing this issue and missed values imputation was executed through ML techniques which leads to higher performance and classification or estimation precision (9). Feature selection (FS) was also executed in this study, but certain significant and information rich features were ignored at the time of model development. Furthermore, statistical techniques will be utilized for model generation which leads to poor predictive performance (10). An intelligent approach is employed in order to find solutions to the existing issues and provide churn prediction with the highest possible degree of precision. This article develops a novel Tomtit Flock Optimization with Deep Variational Autoencoder for Customer Churn Prediction (TFODVAE-CCP) in the telecommunication industry. The goal of the presented TFODVAE-CCP technique mainly concentrates on the identification and classification of customer churns. Firstly, the presented TFODVAE-CCP technique performs data pre-processing at the primary stage. Next, the TFODVAE-CCP technique employs TFO algorithm to elect an optimal subset of features and thereby improve classification accuracy. Finally, the DVAE method will be leveraged for classifying churner/non-churner. A

detailed set of experimental analyses is made on telecommunication churn dataset. From Beeharay and Tsokizep Fokone (11), suggested that a two-layer flexible voting ensemble method for predicting the CC rates in telecom sector. The data leveraged in this work are from Duke University and IBM Sample Data Sets. Following the pre-processing phase, the data were classified into an imbalanced as well as balanced set. The balanced set has an equivalent number of examples for both classes ('not churn' and 'churn'). Dalli (12) examined numerous published articles that utilized ML approaches. However, it can be noticed that an incredible lack of empirically derived heuristic data in which one should affect the hyper parameters accordingly. The purpose of the study carried out by Liyanage *et al.*, (13) was to investigate data on about 7000 post-paid customers by taking into account 21 different characteristics. First, the information has been processed using machine learning techniques such as KNN, ANN, and others. In addition to this, DNNs have identified other hidden layers in the data. The long short-term memory networks (LSTM) can be deemed under the DNN to produce outcomes better than the other methods. Ullah *et al.*, (14) devise a churn predictive technique that uses classification along with the clustering approaches for identifying the churn users and offers the aspects behind the CC in the telecommunication sector. FS can be conducted by means of information gain and relation attribute ranking filters. This modeled strategy will classify churn customers' data applying classification ways where the RF method worked well, with 88.63% of the samples being appropriately categorized. Creating an effective retaining policy becomes an essential task for customer relationship management (CRM) in order to reduce churn. Once categorization is done, this devised method will segment churning customers' facts through categorization of churn customers in crews utilizing cosine similarity for offering group-related retention offers. Vijaya and Sivasankar (15) modeled a technique that might be used to determine which characteristics of telecom CCP were the most effective. Then the selective features are given to the ensemble-classifier approaches namely Random Subspace, Bagging, and Boosting. In this study, duke varsity - churn prediction dataset is under consideration

for evaluating performance and three sets of experiments were executed. Shama *et al.*, (16) devise XGBoost technique as a method having optimal performance among other existing techniques. The previously utilized techniques focus more on the precise estimation of churners than non-churners, where this devised technique will classify churners among the total churners properly and reach the highest True positive rates. Hooda and Mittal (17) suggested an Optimized Kernel MSVM classification method was modeled for predicting and classifying churn. In this modeled work, MSVM technique was utilized for classification. The kernel PCA and ALO optimizer approach were utilized for feature extraction and selection (17). Mohammad *et al.* (18), explores how machine learning can be used to predict when customers are likely to leave their telecommunication service provider. The authors focus on the identification and classification of customer churns using a variety of machine learning classifiers. Caigny *et al.* (19) test their approaches using real-world data from the telecommunication industry and compare their results with other existing models. Ultimately, their findings suggest that machine learning classifiers can be a useful tool for telecommunication companies to predict customer churn and retain their customers.

The Proposed Model

In this article, a new TFODVAE-CCP technique was devised for customer churn classification in the telecommunication sector. The presented TFODVAE-CCP technique mainly concentrates on the identification and classification of customer churns. The presented TFODVAE-CCP technique follows a three-stage process: data pre-

processing feature selection process using Tomtit flock optimization, classification process using deep variational autoencoder model. Figure 1 demonstrates the overall process of TFODVAE-CCP methodology.

Data Normalization

The min-max method is a simple and familiar normalization method in data analysis. Using this technique, the data will be scaled to a uniform level and the data changing edges can be dispersed in the interval of 0 and 1. With the assumption of the attribute X , it can map the dataset among X_{min} and X_{max} , the min-max normalization (X_{norm}) can be defined as follows (20).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}. \quad [1]$$

Algorithmic Steps of TFO based Feature Selection Technique

At this stage, the TFODVAE-CCP technique employed the TFO algorithm (21) to elect an optimal subset of features and thereby improve classification accuracy. The TFO approach of imitating the performance of flocks of the tomtits is a hybrid system to determine the global conditional extreme of the functions of several variables, compared with either SI or bio-inspired techniques. A tomtit unites in flocks. It can follow the commands of flock leaders and have several freedoms in selecting the approach for seeking food. Tomtits are notable from other birds and animals by their special cohesion, coordination of combined actions, and intensity of utilization of the food source establish still it disappears, and obvious and friendly implementation of commands common to members of flocks.

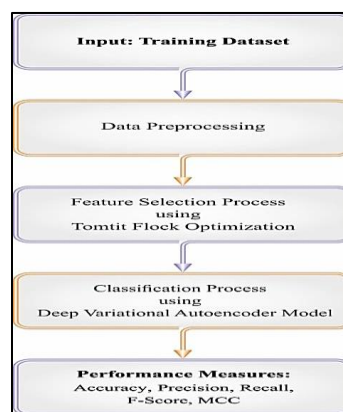


Figure 1: Overall process of TFODVAE-CCP system

Finite sets $l = \{x^j = (x_1^j, x_2^j, \dots, x_n^j)^T, j = 1, 2, \dots, NP\} \subset D$ of probable solution termed as populations were utilized for solving this problem of determining a global constrained minimal of objective function (OF), whereas x^j denotes the tomtit-individual (possible solution) with number j , and NP denotes the population size.

Initially, if the count of iterations $k = 0$, this technique generates primary population of tomtits using the uniform distribution law on possible solutions set D . The value of OFs $f(x)$ has been computed to all the tomtits that are feasible solutions. The solution with optimum value of OFs is the position of flock leader. The leader could not be searching for food in the present iteration. It delays for outcomes of determining food to process outcomes and more storing from the memory matrix (14):

$$\{f(x^1) : f(x^K)\} \quad [2]$$

The memory matrix size is $\times (n + 1)$, whereas K is a provided maximal count of records. An optimum attained outcome on every k^{th} iteration was further in the matrix still the matrix was satisfied. The number K determines the iteration count in one pass. The outcomes in the matrix are arranged by the subsequent rule. The first record in the matrix is an optimum solution $(x^1, f(x^1))$, another recorded were orderly by enhancing (non-decreasing) value of OFs. Once the memory matrix was satisfied, an optimum outcome was located in a different set Pool (group of optimum outcomes of passes), afterward that the matrix was empty.

A novel place of the leader was arbitrarily created by Levy distribution (15):

$$x_i^{1,k+1} = x_i^{1,k} + \frac{\alpha}{k+1} \cdot \text{Levy}_i(\lambda), i = 1, n, \quad [3]$$

Whereas $x_i^{1,k}$ implies the co-ordinate of leader positions on k^{th} iteration and α stand for the movement step, $\lambda \in [1, 3]$. Research on animal performance is demonstrated that the Levy distribution accurately possible refers to the trajectory of birds and insects. This procedure explains the flight of the flock leader around. Another tomtit flies afterward at the command of leaders. The positions of this tomtit are modeled utilizing a uniform distribution on the parallelepiped set. It can be considered that every j^{th} individual is a memory that is stored:

- The current iteration count k ;
- The current position $x^{j,k}$ and the equivalent value of OFs $f(x^{j,k})$;
- An optimum position $x^{j,best}$ and the equivalent value of OFs from the population $f(x^{j,best})$,
- An optimum position of tomtits $x^{j,best}$ in every iteration and the equivalent value of OFs $f(x^{j,best})$,
- An optimum position $x^{j,local}$ amongst every tomtit placed in the vicinity of j^{th} individual of radius p and the equivalent value of OFs $f(x^{j,local})$.

The trajectory of movement of all the individuals (for all tomtits $j = 1, \dots, NP$) on the segmented $[0; T_t]$, in that the searching was implemented at the present iteration was explained by solution of the stochastic differential formula:

$$\begin{aligned} dx^{j,k} &= f(x^{j,k}(t)) dt + \sigma(x^{j,k}(t)) dW + dq, x^{j,k}(0) \\ &= x^{j,k}, j \\ &= 2, \dots, NP, \end{aligned} \quad [4]$$

Whereas $W(t)$ defines the typical Wiener stochastic procedure, T_t refers to the assigned time by leader of packs to search for members of packs at the present iteration, dq denotes the Poisson element that is formulated as:

$$dq = \sum_p \delta(t - \tau_p) \theta_p, \quad [5]$$

$\delta(t)$ Signifies the asymmetric delta function, τ_p represents the moments of jumps. In random moments of times τ_p the place of tomtit experiences arbitrary increments θ_p , establishing the Poisson stream of events of provided intensity μ . The solution of formula defines the trajectories of tomtit's movement which execute the diffusion searching process with jumps.

The drift vector $f(x^{j,k}(t)) \forall t \in [0; T]$ was explained by the formula (16):

$$\begin{aligned} f(x^{j,k}(t)) &= c_1 r_1 [x^{j,best} \\ &- x^{j,k}(t), \end{aligned} \quad [6]$$

Particularly, it proceeds into account data on optimum solutions from the populations: the position of the global leader of flock. The diffusion matrix $\sigma(x^{j,k}(t))$ proceeds into account data on

optimum solution achieved by provided separate to every past iteration, about optimum solution from the vicinity of present solution, defined by radius p :

$$\begin{aligned} \sigma(x^{j,k}(t)) &= c_2 r_2 [x^{j,best} - x^{j,k}(t)] \\ &+ c_3 r_3 [x^{j,local} - x^{j,k}(t)], \end{aligned} \quad [7]$$

in which c_1, c_2, c_3 demonstrates the outcome of coefficients; r_1, r_2, r_3 stands for the random parameters uniformly distributed on the segment of zero and one. The parameter c_2 defines the procedure of forgetting one's searching history; parameter c_3 defines the leader effect amongst neighbors.

The outcome of stochastic differential equation is an arbitrary procedure, the ways of which has sections of continuous variation; interrupt by jumps of provided intensity. It defines the movement of tomtits, attended by comparatively short jumps. This solution is established by numerical integration with step size h . When one coordinate value hits the boundary of searching region or goes beyond it, afterward it can be obtained equivalent to value on this boundary. An optimum position gained in the present iteration was selected as novel position $X^{j,k,search}$ of tomtits. Amongst each novel position of tomtits, an optimum one was selective. It can be recorded in the memory matrix and recognized with the final place of leaders of the flock at the present iteration, and then the next iteration starts with the process to determine a novel position of leaders of the flock and the primary position of members of flocks relating to it. This technique ends once the maximal count of passes P was obtained.

The fitness function of the TFO technique is derived from having a trade-off among the chosen feature count in all solutions and classification accuracy by the chosen feature set. The fitness function can be defined using Eq. [8]:

$$Fitness = \alpha \gamma_R(D) + \beta \frac{|R|}{|C|} \quad [8]$$

where $\gamma_R(D)$ represents classifier error rate of the DVAE model. $|R|$ denotes cardinality of the elected

subset and $|C|$ is the total number of features in the dataset, α , and β are two parameters representing the significance of classification quality and subset length. $\alpha \in [1,0]$ and $\beta = 1 - \alpha$.

DVAE Based Customer Churn Classification

In this study, the DVAE method will be employed for classifying churners/non-churner. Autoencoder (AE) is a variant of neural networks used for data classification. It encompasses hidden layer h for defining code applied to denote input data (22). The AE includes an encoding unit function $z = f(x)$ and decoding unit function $r = g(z)$, whereas x represents input. The useful attributes from AE with the constraints that z to hold minimal dimension over x . The dimension of the AE was superior to input dimensions. The learning process applies the AE to capture useful features from the trained data. The VAE was a directed network that exploits learned approximate implications and experiences training by the use of gradient models. To generate samples from method, VAE originally derive a sample z from the code distribution $p_{model}(z)$. The samples are carried out via differentiable generator networks $g(z)$. At last, x undergoes sampling from distribution $P_{model}(x; g(z)) = p_{model}(x|z)$. During the training procedure, projected inference network (or encoder) $q(z|x)$ was applied to attain z and $P_{model}(x|z)$ was then reflected as the decoding network. The major view of VAE is that it undergoes training by maximizing variation lower bound $L(q)$ relevant to data instance x :

$$\begin{aligned} L(q) &= E_{z \sim q(x)} \log p_{model}(z, x) \\ &+ H(q(z|x)) \end{aligned} \quad [9]$$

$$\begin{aligned} &= E_{z \sim q(z|x)} \log p_{model}(x|z) \\ &- D_{KL}(q(z|x) || p_{model}(z)) \end{aligned} \quad [10]$$

The first element in Eq. [9] denotes cumulative log likelihood of visible and hidden variables. Later, second element indicates estimated posterior entropy. When q is chosen by Gaussian distribution, noise is added to assessed average value, maximizing entropy element stimulated increase in noise standard deviations. Figure 2 demonstrates the structure of VAE technique.

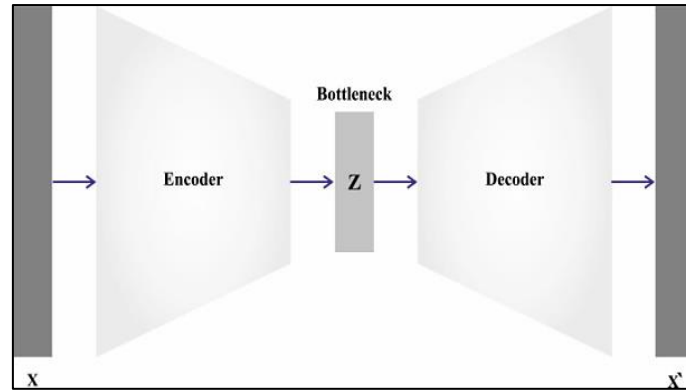


Figure 2: Structure of VAE

The entropy component encourages variational posterior in the placement of high probability mass on numerous z values that produce x , rather than weakening to a particular point approximation. In addition, the first element in Eq. [9] represents reconstruction log-likelihood. Afterward, the second element signifies the projected posterior distribution $q(z|x)$ and model prior $p_{model}(z)$. Stochastic gradient descent (SGD) over BP manages stochastic inputs, then not stochastic units in the network. The solution can be called “reparameterization trick”, for transferring the samples to the input layer. It can be simple from $N(\mu(x), \theta(x))$ by sampling $\epsilon \sim N(0, I)$, after determining $p_{model}z = \mu(x) + \theta^{1/2}(x) * \epsilon$. Here, $\mu(x)$ and $\theta(x)$ imply mean and covariance of $z|x$. Therefore, Eq. [10] can be rewritten in Eq. [11]:

$$L(q) = E_{\epsilon \sim N(0, I)} p_{model}(x|z = \mu(x) + \theta^{1/2}(x) * \epsilon) - D_{KL}(q(z|x) || p_{model}(z)) \quad [11]$$

VAE includes input, different AE, and resultant layers. After that, an unsupervised pre-training stage, and supervised fine-tuning stage were applied to learn the complete variables of network with the help of the BP approach. This approach has one input, five hidden, and one output layer.

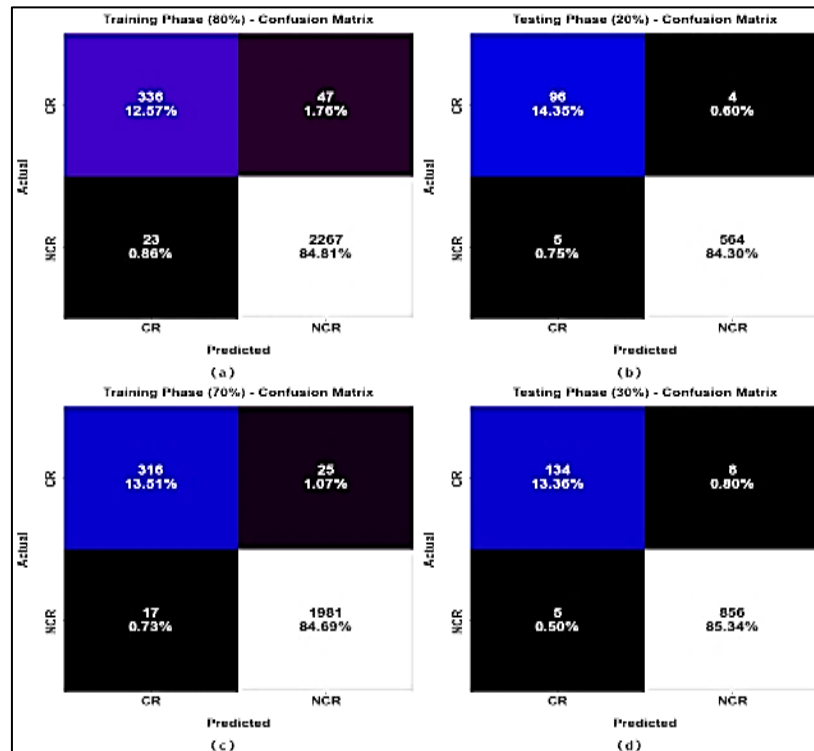
Results and Discussion

In this section, the CCP performance of the TFODVAE-CCP approach is investigated in detail.

The presented TFODVAE-CCP model can be tested on a dataset, comprising 3342 samples with two class labels as demonstrated in Table 1. The confusion matrices offered by the TFODVAE-CCP system on churn dataset in Figure 3. The Figurers portrayed that the TFODVAE-CCP model has accomplished effectual churn classification performance. For instance, on 80% of TR database, the TFODVAE-CCP model has recognized 336 samples into CR class and 2267 samples into NCR class. Also, on 20% of TS database, the TFODVAE-CCP approach has recognized 96 samples into CR class and 564 samples into NCR class. In addition, on 70% of TR database, the TFODVAE-CCP method has detected 316 samples into CR class and 1981 samples into NCR class. Table 2 provides brief CCP outcomes of the TFODVAE-CCP technique on 80% of TR and 20% of TS database. Figure 4 inspects the overall CCP performance of the TFODVAE-CCP model under 80% of TR database. The Figurers implied that the TFODVAE-CCP system has shown enhanced performance on both classes. For instance, on CR class, the TFODVAE-CCP model has offered $accu_y$ of 97.38%, $prec_n$ of 93.59%, $reca_l$ of 87.73%, F_{score} of 90.57%, and MCC of 89.11%. Also, on NCR class, the TFODVAE-CCP technique has rendered $accu_y$ of 97.38%, $prec_n$ of 97.97%, $reca_l$ of 99%, F_{score} of 98.48%, and MCC of 89.11%.

Table 1: Dataset Details

Label	Class	No. of Samples
CR	Churners	483
NCR	Non-Churner	2859
Total Number of Samples		3342

**Figure 3:** Confusion Matrices of TFODVAE-CCP System
(a-b) TR and TS of 80:20 and (c-d) TR and TS of 70:30**Table 2:** CCP Outcomes Analysis of TFODVAE-CCP System Under 80:20 of TR/TS Database

Labels	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	MCC
Training Phase (80%)					
CR	97.38	93.59	87.73	90.57	89.11
NCR	97.38	97.97	99.00	98.48	89.11
Average	97.38	95.78	93.36	94.52	89.11
Testing Phase (20%)					
CR	98.65	95.05	96.00	95.52	94.73
NCR	98.65	99.30	99.12	99.21	94.73
Average	98.65	97.17	97.56	97.37	94.73

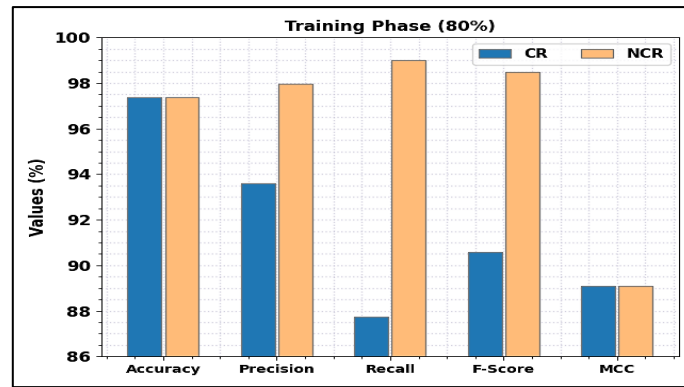


Figure 4: CCP outcomes analysis of TFODVAE-CCP system under 80% of TR database

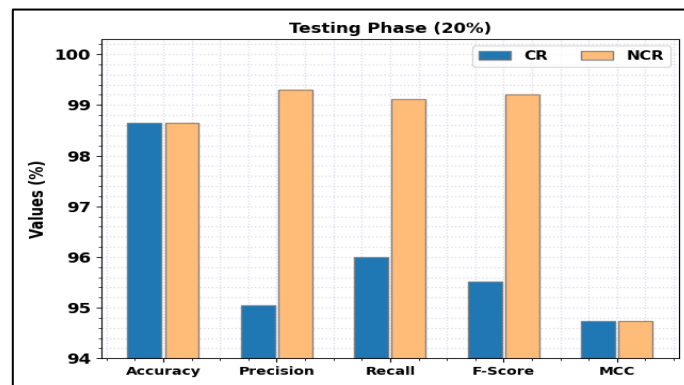


Figure 5: CCP Outcomes Analysis of TFODVAE-CCP System Under 20% of TS Database

Table 3: CCP Outcomes Analysis of TFODVAE-CCP System Under 70:30 of TR/TS Database

Labels	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	MCC
Training Phase (70%)					
CR	98.20	94.89	92.67	93.77	92.73
NCR	98.20	98.75	99.15	98.95	92.73
Average	98.20	96.82	95.91	96.36	92.73
Testing Phase (30%)					
CR	98.70	96.40	94.37	95.37	94.63
NCR	98.70	99.07	99.42	99.25	94.63
Average	98.70	97.74	96.89	97.31	94.63

Figure 5 examines the overall CCP performance of the TFODVAE-CCP approach under 20% of TS database. The Figure denoted the TFODVAE-CCP approach has exhibited enhanced performance on both classes. For example, on CR class, the TFODVAE-CCP algorithm has offered $accu_y$ of 98.65%, $prec_n$ of 95.05%, $reca_l$ of 96%, F_{score} of 95.52%, and MCC of 94.73%. Similarly, on NCR class, the TFODVAE-CCP approach has presented $accu_y$ of 98.65%, $prec_n$ of 99.30%, $reca_l$ of

99.12%, F_{score} of 99.21%, and MCC of 94.73%. Table 3 delivers brief CCP outcomes of the TFODVAE-CCP technique on 70% of TR and 30% of TS database. Figure 6 examines the overall CCP performance of the TFODVAE-CCP algorithm under 70% of TR database. The Figure denoted the TFODVAE-CCP approach has exhibited enhanced performance on both classes. For example, on CR class, the TFODVAE-CCP method has presented $accu_y$ of 98.20%, $prec_n$

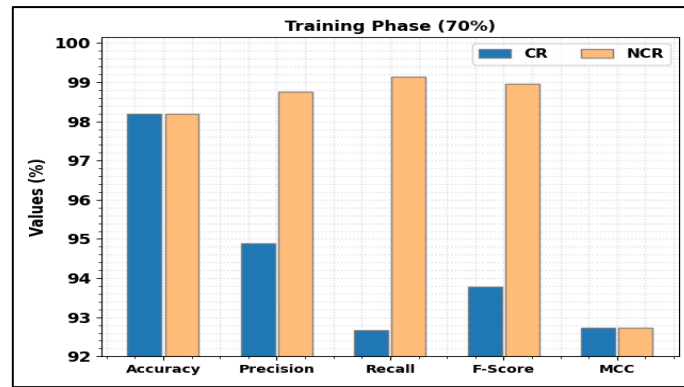


Figure 6: CCP Outcomes Analysis of TFODVAE-CCP System Under 70% of TR Database

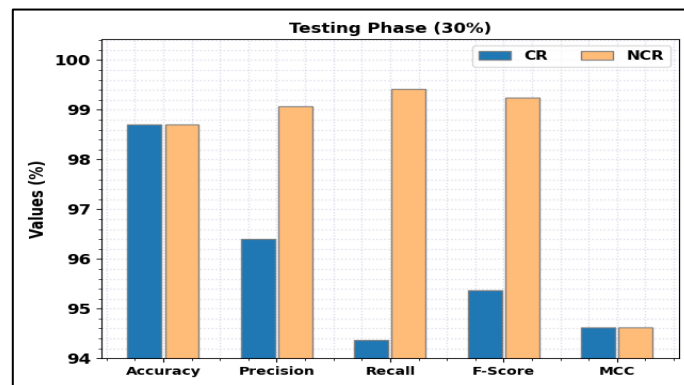


Figure 7: CCP Outcomes Analysis of TFODVAE-CCP System Under 30% of TS Database

of 94.89%, $reca_l$ of 92.67%, F_{score} of 93.77%, and MCC of 92.73%. Similarly, on NCR class, the TFODVAE-CCP methodology has presented $accu_y$ of 98.20%, $prec_n$ of 98.75%, $reca_l$ of 99.15%, F_{score} of 98.95%, and MCC of 92.73%. Figure 7 scrutinizes the overall CCP performance of the TFODVAE-CCP method under 30% of TS database. The figure represented the TFODVAE-CCP method has displayed enhanced performance on both classes. For example, on CR class, the TFODVAE-CCP method has rendered $accu_y$ of 98.70%, $prec_n$ of 96.40%, $reca_l$ of 94.37%, F_{score} of 95.37%, and MCC of 94.63%. Likewise, on NCR class, the TFODVAE-CCP method has presented $accu_y$ of 98.70%, $prec_n$ of 99.07%, $reca_l$ of 99.42%, F_{score} of 99.25%, and MCC of 94.63%.

The training accuracy (TRA) and validation accuracy (VLA) acquired by the TFODVAE-CCP

approach under test database is exemplified in Figure 8. The training accuracy for 25 epochs has been attained by 98.70%. The validation accuracy for 25 epochs has been attained by 97.30%. The simulation result denoted the TFODVAE-CCP approach has attained maximal values of TRA and VLA. Seemingly the VLA is greater than TRA. The training loss (TRL) and validation loss (VLL) obtained by the TFODVAE-CCP algorithm under test database are shown in Figure 9. In Figure 9 the training loss for 25 epochs has been attained by 0.070 and the validation loss for 25 epochs has been attained by 0.68. The simulation result showed the TFODVAE-CCP algorithm has exemplified least values of TRL and VLL. Particularly, the VLL is lesser than TRL.

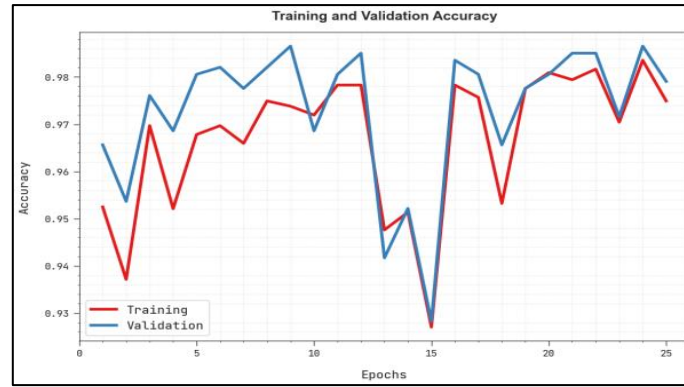


Figure 8: TRA and VLA Analysis of TFODVAE-CCP System

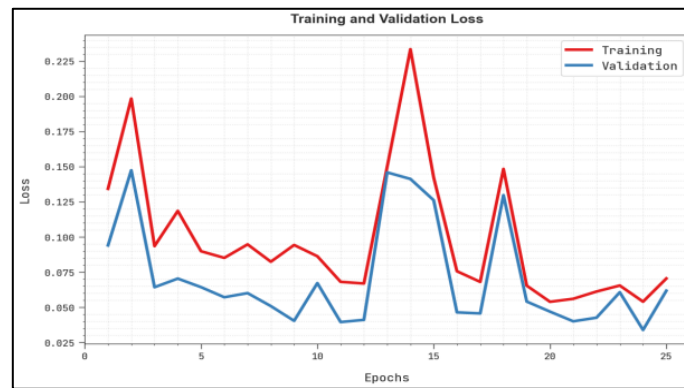


Figure 9: TRL and VLL Analysis of TFODVAE-CCP System

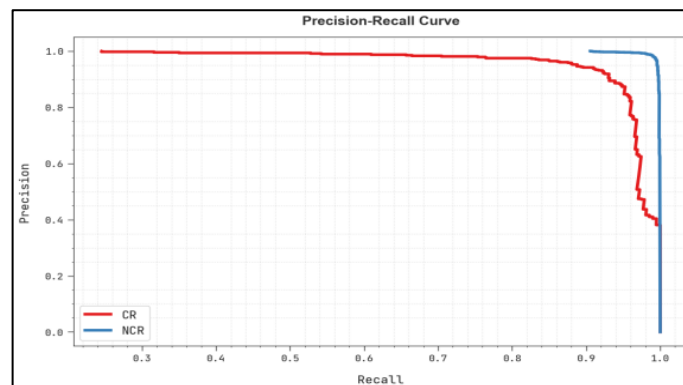


Figure 10: Precision-Recall Analysis of TFODVAE-CCP System

A clear precision-recall investigation of the TFODVAE-CCP algorithm under test database is represented in Figure 10. The precision and recall for the proposed model are 97.74 and 96.89 respectively. The Figure exemplified the TFODVAE-CCP technique has resulted in enhanced values of precision-recall values in every class label. Table 4 provides a detailed comparative study of the TFODVAE-CCP approach with other existing techniques suggested by Haridasan (23). Figure 11 inspects a comparative $accu_y$ investigation of the TFODVAE-CCP system

with other techniques. The Figure represented that the XGBoost, RF, and Adaboost models have offered decreased $accu_y$ of 80.89%, 81.47%, and 81.89% respectively. Followed by, the PCPM model has shown reduced $accu_y$ of 86.51% whereas the OWELM and WELM models have demonstrated reasonable $accu_y$ of 88.9% and 89.82% respectively. Though the AOA-SBLSTM model has resulted in reasonable $accu_y$ of 96.21%, the TFODVAE-CCP model outperformed compared methods with higher $accu_y$ of 98.7%.

Table 4: Comparative Analysis of TFODVAE-CCP System with Other Algorithms

Methods	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}
Adaboost Model	81.89	81.35	80.38	80.26
RF Model	81.47	80.90	80.41	80.39
XGBoost Model	80.89	80.75	80.54	78.81
OWELM Model	88.90	91.67	87.71	87.24
WELM Model	89.82	89.29	85.36	89.85
PCPM Model	86.51	89.44	87.02	87.04
AOA-SBLSTM Model	96.21	96.80	88.44	92.07
TFODVAE-CCP	98.70	97.74	96.89	97.31

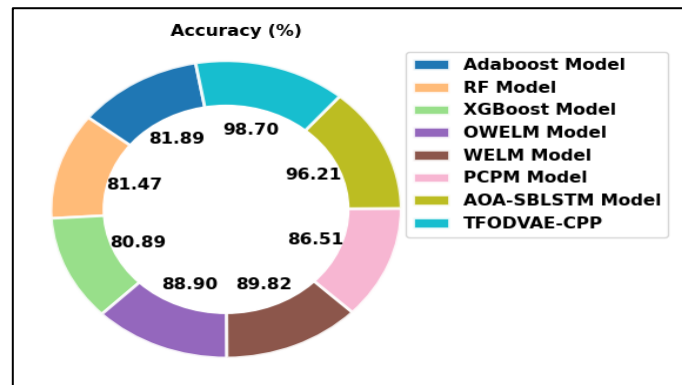
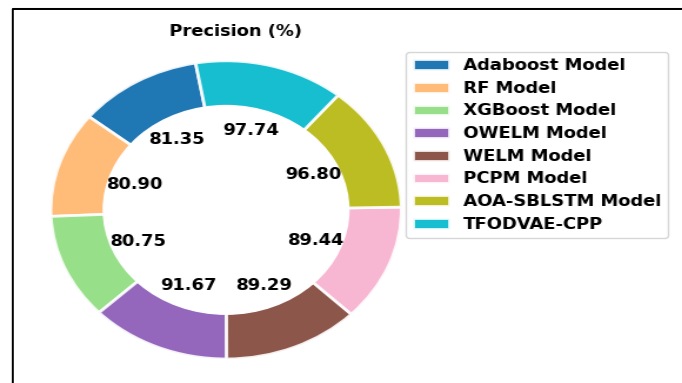
**Figure 11:** $Accu_y$ Analysis of TFODVAE-CCP System with Other Algorithms**Figure 12:** $Prec_n$ Analysis of TFODVAE-CCP System with Other Algorithms

Figure 12 reviews a brief $prec_n$ inspection of the TFODVAE-CCP approach with other methods. The Figure denoted the XGBoost, RF, and Adaboost approaches have rendered decreased $prec_n$ of 80.75%, 80.90%, and 81.35% correspondingly. Then, the PCPM technique displayed reduced $prec_n$ of 89.44% where the OWELM and WELM

techniques have established reasonable $prec_n$ of 91.67% and 89.29% correspondingly. Though the AOA-SBLSTM method has resulted in reasonable $prec_n$ of 96.80%, the TFODVAE-CCP algorithm displayed compared techniques with higher $prec_n$ of 97.74%.

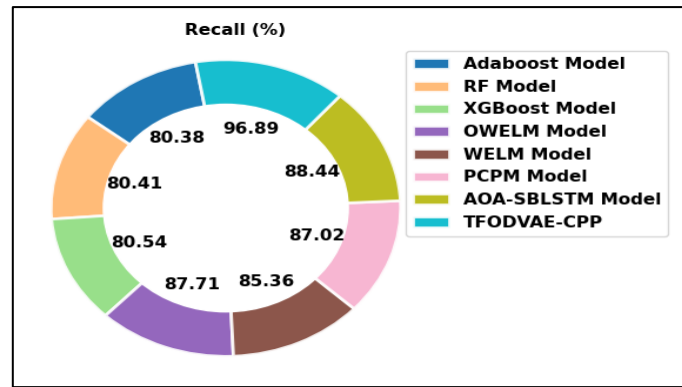


Figure 13: $Recall_i$ Analysis of TFODVAE-CCP System with Other Algorithms

Figure 13 examines a detailed $recall_i$ inspection of the TFODVAE-CCP method with other methodologies. The Figure denoted the XGBoost, RF, and Adaboost methods presented decreased $recall_i$ of 80.54%, 80.41%, and 80.38% correspondingly. Then, the PCPM approach exhibited reduced $recall_i$ of 87.02% whereas the OWELM and WELM models have demonstrated reasonable $recall_i$ of 87.71% and 87.02% correspondingly. Though the AOA-SBLSTM approach has resulted in reasonable $recall_i$ of 88.44%, the TFODVAE-CCP method outperformed compared methods with higher $recall_i$ of 96.89%. Figure 14 scrutinizes a comparative F_{score} analysis of the TFODVAE-CCP algorithm with other

methods. The Figure denoted the XGBoost, RF, and Adaboost techniques have presented decreased F_{score} of 78.81%, 80.39%, and 80.26% correspondingly. Then, the PCPM approach has shown reduced F_{score} of 87.04% whereas the OWELM and WELM methodologies have demonstrated reasonable F_{score} of 87.24% and 89.85% correspondingly. Though the AOA-SBLSTM approach has resulted in reasonable F_{score} of 92.07%, the TFODVAE-CCP method outperformed compared techniques with increased F_{score} of 97.31%. Therefore, the presented TFODVAE-CCP model has reported maximum CCP process.

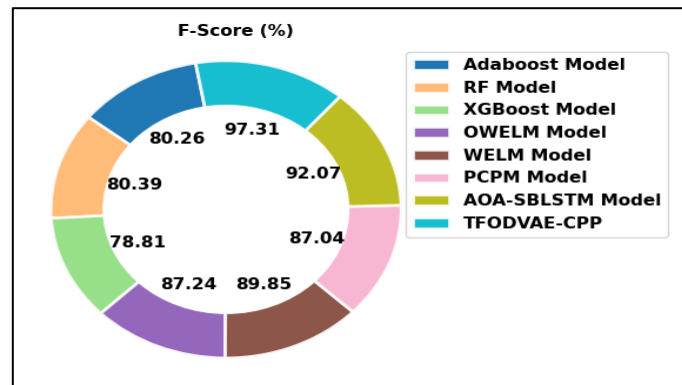


Figure 14: F_{score} Analysis of TFODVAE-CCP System with Other Algorithms

Conclusion

The TFODVAE-CCP approach presents a promising solution for accurate and timely customer churn classification in the telecommunication sector. The three-stage process of data pre-processing, TFO based FS, and VAE classification effectively identifies and classifies customer churns, resulting in improved classification accuracy. The experimental analyses

conducted on telecommunication churn data demonstrate the superiority of the TFODVAE-CCP method compared to other existing models. Future work can focus on hyper parameter tuning of the DVAE model using hybrid metaheuristics to further enhance the performance of the approach. The specific measures of our proposed model achieves the accuracy, precision, recall and F1-score has been attained by 98.70%, 97.74%, 96.89% and 97.31%. Overall, the TFODVAE-

CCP technique can be a valuable tool for telecommunication companies to prevent customer churn and retain their customer base.

Abbreviation

Nil.

Acknowledgement

Nil.

Author Contributions

Nil.

Conflict of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Ethics Approval

Nil.

Funding

The authors received no specific funding for this study.

References

1. Pustokhina IV, Pustokhin DA, Nguyen PT, Elhoseny M, Shankar K. Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector. *Complex & Intelligent Systems*. 2021 Apr 5; 9: 3473–3485.
2. Lalwani P, Mishra MK, Chadha JS, Sethi P. Customer churn prediction system: a machine learning approach. *Computing*. 2022 Feb; 104(2): 271-94.
3. Fujo SW, Subramanian S, Khder MA. Customer churns prediction in telecommunication industry using deep learning. *Information Sciences Letters*. 2022 Jan; 11(1): 185-198.
4. Wu S, Yau WC, Ong TS, Chong SC. Integrated churn prediction and customer segmentation framework for telco business. *IEEE Access*. 2021 Apr 16; 9: 62118-36.
5. Kanwal S, Rashid J, Kim J, Nisar MW, Hussain A, Batool S, Kanwal R. An attribute weight estimation using particle swarm optimization and machine learning approaches for customer churn prediction. In *2021 International Conference on Innovative Computing (ICIC)* 2021 Nov 9; (pp. 1-6): IEEE.
6. Jain H, Yadav G, Manoov R. Churn prediction and retention in banking, telecom and IT sectors using machine learning techniques. In *Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019 2020 Jul 26* (pp. 137-156). Singapore: Springer Singapore.
7. Jain H, Khunteta A, Srivastava S. Telecom churn prediction and used techniques, datasets and performance measures: a review. *Telecommunication Systems*. 2021 Apr; 76: 613-30.
8. Edwine N, Wang W, Song W, Ssebuggwawo D. Detecting the risk of customer churn in telecom sector: a comparative study. *Mathematical Problems in Engineering*. 2022; 2022(1): 8534739.
9. Majhi B, Rajput SS, Majhi R. Performance Evaluation of Machine Learning Techniques for Customer Churn Prediction in Telecommunication Sector. In *Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science*. IGI Global. 2021; pp. 262-274.
10. Xu T, Ma Y, Kim K. Telecom churn prediction system based on ensemble learning using feature grouping. *Applied Sciences*. 2021 May 21; 11(11): 4742.
11. Khalaj M, Taghizadeh Harat A. Improved Ensemble Learning Model by Swarm Intelligence for Mobile Subscribers' Churn Prediction. *Signal and Data Processing*. 2024 Mar 10; 20(4): 45-66.
12. Dalli A. Impact of hyperparameters on deep learning model for customer churns prediction in telecommunication sector. *Mathematical Problems in Engineering*. 2022; 2022(1): 4720539.
13. Liyanage RP, Kumara BT, Kuhaneswaran B, Prasanth S. Deep Learning Approach for Detecting Customer Churn in Telecommunication Industry. In *Social Customer Relationship Management (Social-CRM) in the Era of Web 4.0*. IGI Global. 2022; pp. 196-215.
14. Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*. 2019 May 6; 7: 60134-49.
15. Vijaya J, Sivasankar E. Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector. *Computing*. 2018 Aug; 100: 839-60.
16. Sharma T, Gupta P, Nigam V, Goel M. Customer churn prediction in telecommunications using gradient boosted trees. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2019*. Springer Singapore. 2020; Volume 2: pp. 235-246.
17. Hooda P, Mittal P. An Optimized Kernel MSVM Machine Learning-based Model for Churn Analysis. *International Journal of Advanced Computer Science and Applications*. 2022; 13(5):487-494.
18. Mohammad NI, Ismail SA, Kama MN, Yusop OM, Azmi A. Customer churn prediction in telecommunication industry using machine learning classifiers. In *Proceedings of the 3rd international conference on vision, image and signal processing 2019 Aug 26*; pp. 1-7.
19. De Caigny A, Coussement K, De Bock KW. A new hybrid classification algorithm for customer churns prediction based on logistic regression and decision trees. *European Journal of Operational Research*. 2018 Sep 1; 269(2):760-72.
20. Kiran A, Vasumathi D. Data mining: min-max normalization based data perturbation technique for privacy preservation. In *Proceedings of the third international conference on computational*

- intelligence and informatics: ICCII 2018. Springer Singapore. 2020 Mar 18;pp. 723-734.
21. Panteleev AV, Kolessa AA. Application of the tomtit flock metaheuristic optimization algorithm to the optimal discrete time deterministic dynamical control problem. *Algorithms*. 2022 Aug 26; 15(9): 301.
 22. Kingma DP, Welling M. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*. 2019 Nov 27; 12(4): 307-92.
 23. Haridasan V, Muthukumaran K, Hariharanath K. Arithmetic Optimization with Deep Learning Enabled Churn Prediction Model for Telecommunication Industries. *Intelligent Automation & Soft Computing*. 2023 Mar 1; 35(3):3531-3544.