Intelligent Optimal Deep Learning based Customer Churn Prediction Model in Telecom Industry

Taif Khalid Shakir[1], Dr. Ahmed N. Al Masri[2,*]

[1] Ascencia Business School, College de Paris, France

[2] American University in the Emirates, Dubai, United Arab Emirates

Emails: taif.shakir@cabling.att-mail.com ; ahmed.almasri@aue.ae

Abstract

Customer churn prediction (CCP) is a crucial problem in telecom industry which helps to improve the revenue of the company and prevent the loss of customers. Customer churn is an important issue in service sector with highly competitive services. At the same time, the prediction of users who are probably leaving the company can be identified at an earlier stage to prevent loss of revenue. Several works have used machine learning (ML) techniques for predicting the existence of customer churn in different industries. With this motivation, this paper presents an optimal long, short-term memory with stacked autoencoder (OLSTM-SAE) technique for CCP in telecom industry. The OLSTM-SAE technique encompasses three subprocesses namely preprocessing, classification, and parameter optimization. The OLSTM-SAE technique classifies the preprocessed data into churn and non-churn customers. In addition, the grey wolf optimization (GWO) technique is used to adjust the variables involved in the LSTM-SAE model. For examining the enhanced performance of the OLSTM-SAE technique, an extensive simulation analysis takes place, and the outcomes are inspected with respect to various measures. The experimental results highlighted the betterment of the OLSTM-SAE technique in terms of different evaluation parameters.

Keywords: Customer churn, Prediction model, Machine learning, Deep learning, Intelligent models.

1. Introduction

Recent technologies have empowered information driven ventures to examine data. Data mining procedures provide the estimation of particular future conduct of customers [1]. The main issue that diminishes the benefit of an organization is customer churn, which is otherwise called customer steady loss or customer turnover [2, 3]. Customer churn can likewise be characterized as the business knowledge interaction of discovering customers that are going to change from a business to its rival [4-6]. In today's enterprises, plenty of options are available which assist customers with getting benefit of a profoundly serious market. One can pick a specialist co-op that offers preferable assistance over others. Accordingly, the profitmaking associations which contend in immersed markets like banks, telecom, and network access organizations, and protection firms unequivocally centered more around keeping current customers than getting new customers [7]. Besides, keeping up with present customers is demonstrated to be significantly less costly than gaining new customers [8].

DOI: 10.5281/zenodo.5502106
Received: May 01, 2021 Accepted: August 29, 2021

To keep their customers, organizations need to have a profound comprehension of why churn occurs. There are a few motivations to be tended to, like disappointment from the organization, cutthroat costs of different organizations, movement of the customers, and customers' requirement for superior assistance [9]. To provide the significance of customers and the greater expenses of drawing in novel customers contrasted and holding existing ones, banks and other customer-subordinate enterprises should have the option to mechanize the way toward anticipating the practices of their customers utilizing customers' information in their data set. Thus, banks couldn't want anything more than to recognize customers with the most elevated probabilities to withdraw from their administrations. Churn expectation empowers the utilization of customers' exchange profiles to decide the probability of a customer leaving a help [10].

Among the past examinations for churn investigation, a widely employed technique is artificial neural networks (ANNs). To adjust the created models, few methods were explored with ANNs, for example, building medium-sized ANN models which were seen to be effective in numerous areas [11]. These examinations show that an assortment of ANN approaches can be applied to expand forecast precision of customer churn. Truth be told, the utilization of ANN in churn forecast has a major resource in regard to different strategies utilized, in light of the fact that the probability of every arrangement made cannot really settle. Recently, deep neural network (DNNs) has been utilized for churn forecast, yet the way toward choosing the preparation hyperparameters for churn displaying requires additional time and exertion, which could make the cycle more trying for experts and scientists [12]. Barely any scientists have zeroed in on deciding the impacts that distinctive hyperparameters have on the exhibition of DNNs during churn expectation [13]. Because of this hole, experimentally determined heuristic information that can direct the determination of hyperparameters when DNNs are utilized for churn demonstrating is as yet inadequate

This paper presents an optimal long short term memory with stacked autoencoder (OLSTM-SAE) technique for CCP in telecom industry. The OLSTM-SAE technique encompasses three subprocesses namely preprocessing, classification, and parameter optimization. The OLSTM-SAE technique classifies the preprocessed data into churn and non-churn customers. In addition, the grey wolf optimization (GWO) algorithm is used to adjust the variables involved in the LSTM-SAE model. For examining the enhanced performance of the OLSTM-SAE technique, an extensive simulation analysis takes place and the results are inspected interms of different measures.

2. Literature Review

Cenggoro et al. [14] established the reasonable method with the help of vector embedding in Dl method. They show that the method could disclose churning users which might be transformed to utilize the prior telecommunication services. The produced vector is very discriminated among the loyal users and churning enables the proposed method for more predicting to determine either a customer will cease her or his service subscriptions or not. Pustokhina et al. [15] design a dynamic CCPBI-TAMO approach. Additionally, the CPIO-FS techniques are used to FS method and reduce computational overhead. As well, LSTM and SAE models are used for classifying the feature decreased data.

Domingos et al. [16] present an experimental analysis of the effect of distinct hyper-parameters while DNN is employed for churn predictions in the banking field. The result from 3 researches shown that the DNN method implemented well compared to MLP while a rectifier function has been employed to activation in the hidden layer and a sigmoid function has been employed in the output layer. Tariq et al. [17] aim are to observe the user activities and execute problem solving consequently. The presented method utilizes the 2D CNN; a method of DL. The presented method is a layered structure which includes 2 distinct stages which are pre-processing and data load and 2D CNN layers. Also, the distributed architecture and Apache Spark parallel are employed for processing the information in parallel environments. Training information is taken from Kaggle through Telco Customer Churn. De Caigny et al. [18] investigate the values included by integrating textual information to CCP methods. It extends the prior study by CNN towards present practice to analyze textual information in CCP, also, with realtime information from European financial service providers, validate an architecture which explains how textual information could be integrated with a prediction method.

DOI: 10.5281/zenodo.5502106
Received: May 01, 2021 Accepted: August 29, 2021

3. The Proposed Model

This paper has presented an effective OLSTM-SAE technique for CCP in telecom industry. The proposed OLSTM-SAE technique involves a 2-stage process namely LSTM-SAE based prediction and GWO based parameter optimization. Initially, the data gathered from the business sectors were pre-processed for converting to useful information [15, 19]. For this purpose, 3 sub procedures are implemented i.e. min-max normalization, data transformation, and class labelling, to improve the quality of business data. The data transformation procedure comprises the transformation of categorical value to arithmetical value. Later, the data instance is assigned to appropriate class label in the class labeling procedure. Lastly, the min-max normalization procedure gets performed for normalizing the information to a uniform standard. The detailed operation of these processes is offered in the following areas.

3.1 Process involved in LSTM-SAE Model

When the information is preprocessed, the classification method is performed using LSTM-SAE models for determining the incidence of non-churners or churners. LSTM is used for classifying the presence of churn. SAE provides raw information using more promising and detailed feature data, i.e., employed for training classifiers using certain contexts, and finds greater performances compared to training the raw information. Now, the encoder functions map a sequence of fundamental characteristics (viz., time series of multivariate inputs) to a set length vectors of new features (viz., latent space), i.e., converted to the related input sequence with the decoder functions [20]. Fig. 1 illustrates the architecture of LSTM. This is able to mine longer and shorter distance dependency using a sequence of fundamental characteristics and is denoted as powerful and compact representation. This is given as follows

$$f_z = \sigma\left(W_f[h_{z-1}, x_z] + b_f\right) \quad (1)$$

$$i_z = \sigma(W_i[h_{z-1}, x_z] + b_i) \quad (2)$$

$$\hat{C}_z = \tanh(W_c[h_{z-1}, x_z] + b_c) \quad (3)$$

$$o_z = \sigma(W_o[h_{z-1}, x_z] + b_o) \quad (4)$$

$$C_z = f_z \otimes C_{z-1} + i_z \otimes \hat{C}_z \quad (5)$$

$$h_z = o_z \otimes \tanh(C_z) \quad (6)$$

Whereas $+$ & $\otimes$ represents the elementwise addition & product, respectively. The matrics W_f, W_i, W_c, W_o signifies linear transformations. C_t represents cell memory state at time z. h_z indicate output at time z.

Where $X^Z = \left(x^{(1)}, x^{(2)}, \dots x^{(Z)}\right)$ represents Z length inputs sequence of fundamental characteristics at different time stamps $z \in \{1, \dots, Z\}$, the S encoding later the Z recursive upgrade Eq. (17) to (22) to make a synthesized output vector (y^Z) of a predefined $r \times 1$ dimensions, demonstrated as follows:

$$y^Z = \Psi^Z\left(x^{(1)}, x^{(2)}, \dots x^{(Z)}\right) \quad (7)$$

using Ψ^Z a non-linear multi-variable vectors valued function collect LSTM cell tasks for Z time stamps. Now y_k^Z indicates kth component of y^Z, afterward replacing y_k^Z to x_k^{AE}, it generates:

$$h_l = \sigma\left(\sum_{k=1}^{r} w_{lk}\, y_k^Z + b_l\right) \quad (8)$$

The aforementioned equations denote extracted feature as a non-linear combination of Z input (viz., Z vectors of fundamental characteristics). This new feature represents a compact representation of the fundamental characteristics in time [21]. For providing concepts of Z fundamental characteristics, it is fixed to make new features and evaluate the Ψ^Z functions to the initial order Taylor polynomial.

DOI: 10.5281/zenodo.5502106
Received: May 01, 2021 Accepted: August 29, 2021

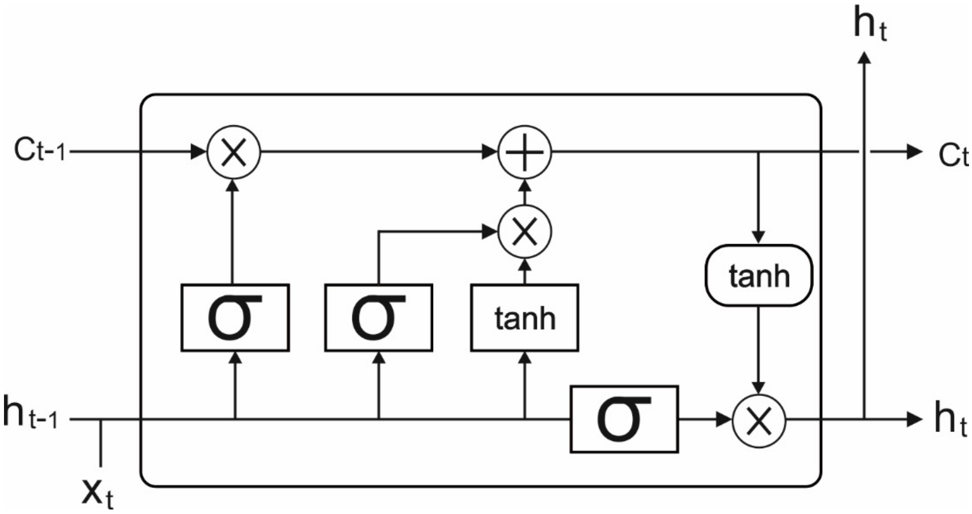

Fig. 1. Structure of LSTM

3.2 Process involved in OLSTM-SAE model

For optimally tune the parameters involved in the LSTM-SAE technique, an optimal parameter tuning process is done by the GWO algorithm. GWO algorithms are presented by Mirjalili et al., [22] in 2016, stimulating in the hunting nature of grey wolf [22-24].

The exploitation phases involve searching for optimum solutions from local search region. The grey wolf attack and encircling the prey in the course of the exploration of optimum solution from local search region. The exploring of prey controls from the exploration phases where the grey wolf search the prey from the global search space. When surrounding the prey, the wolf recognizes the location of prey and gets around them. Now, the location vectors of prey are denoted and the search agent modifies their place regarding the obtained optimal solutions. The surrounded prey is given by:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_p}(k) - \vec{X}(k)\right| \tag{9}$$

$$\vec{X}(k+1) = \overrightarrow{X_p}(k) - \vec{A}.\vec{D} \tag{10}$$

Whereas k denotes the current rounds, $\vec{A}$ & $\vec{C}$ indicates coefficient vector, location vectors of the preys are represented by $\overrightarrow{X_p}(k), \vec{X}$ is the location vector, || means the accurate value, and (.) denotes an elementwise multiplication. The vector $\vec{A}$ and $\vec{C}$ can be given by:

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \tag{11}$$

$$\vec{C} = 2.\vec{r} \tag{12}$$

In the hunting phase, α would direct the action of grey wolves also some contribution is provided to β & δ.

$$\vec{D}_\alpha = |\vec{C}_1 * \vec{X}_\alpha - \vec{X}|$$
$$\vec{D}_\beta = |\vec{C}_2 * \vec{X}_\beta - \vec{X}| \tag{13}$$
$$\vec{D}_\delta = |\vec{C}_1 * \vec{X}_\delta - \vec{X}|$$

Whereas $\vec{D}_\alpha$,$\vec{D}_\beta$, & $\vec{D}_\delta$ indicates the changed distance vectors among the α, β & γ location to other wolfs as well as $\vec{C}_1$,$\vec{C}_2$, & $\vec{C}_3$ denotes 3 coefficient vectors employed in alter distance vectors and it is defined. $\vec{X}$location of vector of other grey wolfs (omega).

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 * \vec{D}_\alpha$$

DOI: 10.5281/zenodo.5502106
Received: May 01, 2021 Accepted: August 29, 2021

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 * \vec{D}_\beta \quad (14)$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 * \vec{D}_\delta$$

whereas $\vec{X}_1$ denotes a reached novel location vector through α location $\vec{X}_\alpha$ and distance vectors $\vec{D}_\alpha$, $\vec{X}_2$ indicate novel location vectors achieved through β location $\vec{X}_\beta$ and distance vectors $\vec{D}_\beta$, $\vec{X}_3$ indicate novel location vector defined through delta location $\vec{X}_\delta$ and distance vectors $\vec{D}_\delta$, and $\vec{A}_1$, $\vec{A}_2$, & $\vec{A}_3$ represents 3 coefficient vectors defined.

$$\vec{X}(k+1) = \frac{\sum_{i=1}^{n} \vec{X}_i}{n} \quad (15)$$

In which $\vec{X}(k+1)$ indicates a novel concluded location vector is defined through the average amount of each location achieved as α, β & δ (n=3).

4. Performance Validation

In order to showcase the improved performance of the OLSTM-SAE technique, a series of simulation analyses take place on the three datasets. Fig. 2 investigates the results analysis of the OLSTM-SAE technique on the dataset-1 interms of accuracy and F-measure. On investigating the results interms of accuracy, it is outperformed that SVM, PCPM, and LDT/UDT techniques have obtained a lower accuracy of 0.789, 0.837, and 0.854.

Table 1 CCP results analysis of OLSTM-SAE model on applied dataset

Methods	Dataset-1		Dataset-2		Dataset-3	
	Accuracy	F-Measure	Accuracy	F-Measure	Accuracy	F-Measure
OLSTM-SAE	0.951	0.949	0.932	0.943	0.927	0.931
LSTM-SAE	0.941	0.936	0.930	0.931	0.914	0.913
ISMOTE-OWELM	0.940	0.935	0.920	0.918	0.909	0.908
SMOTE-OWELM	0.922	0.930	0.918	0.917	0.899	0.899
OWELM	0.906	0.904	0.897	0.894	0.887	0.879
WELM	0.885	0.882	0.876	0.872	0.869	0.852
PCPM	0.837	0.838	0.828	0.831	0.818	0.808
SVM	0.789	0.763	0.725	0.731	0.679	0.682
LDT/UDT	0.854	0.580	0.770	0.670	0.580	0.592

In addition, the WELM and OWELM models have attained slightly enhanced accuracy of 0.885 and 0.906. In line with, the SMOTE-OWELM, ISMOTE-OWELM, and LSTM-SAE techniques have resulted in a moderately closer accuracy of 0.922, 0.94, and 0.951. But the proposed OLSTM-SAE technique has accomplished superior outcomes with an accuracy of 0.941. On examining the results with respect to F-score, it can be stated that LDT/UDT, SVM, and PCPM approached ELM methods have reached somewhat improved F-score of 0.882 and 0.904. Also, the SMOTE-OWELM, ISMOTE-OWELM, and LSTM-SAE techniques have resulted in a moderately closer F-score of 0.930, 0.935, and 0.936. But the presented OLSTM-SAE methodology has accomplished maximum result with an F-score of 0.949.

DOI: 10.5281/zenodo.5502106
Received: May 01, 2021 Accepted: August 29, 2021

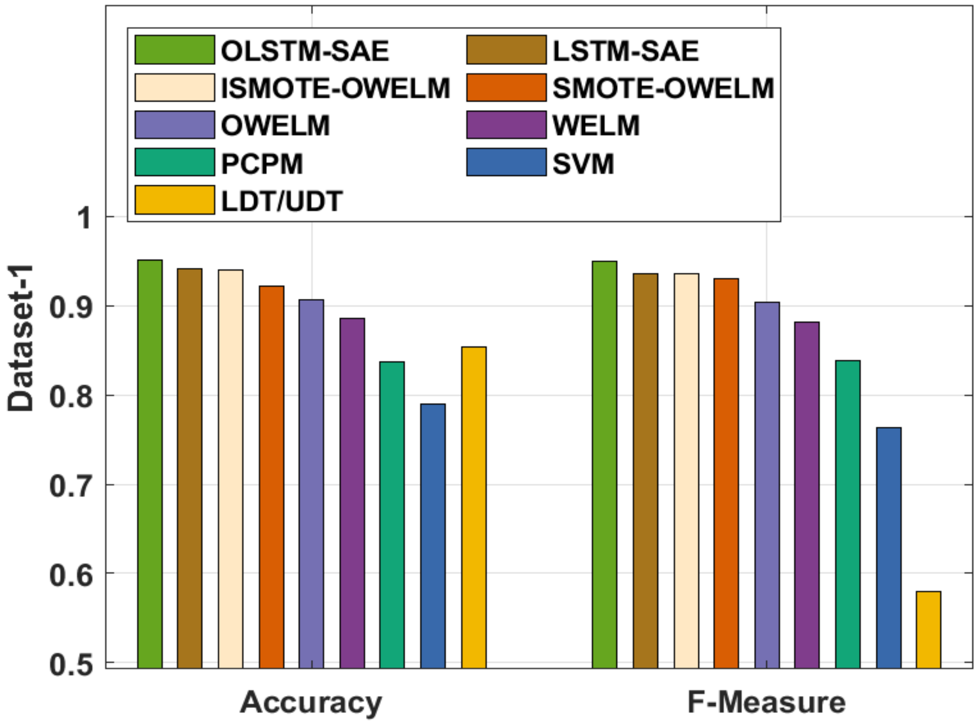

Fig. 2. Comparison study of OLSTM-SAE technique on dataset-1

Fig. 3 examines the outcomes analysis of the OLSTM-SAE manner on dataset-2 with respect to accuracy and F-measure. On inspecting the results with respect to accuracy, it is obvious that SVM, LDT/UDT, and PCPM techniques have obtained a lower accuracy of 0.725, 0.770, and 0.828. Besides, the WELM and OWELM models have attained slightly enhanced accuracy of 0.876 and 0.897. Likewise, the SMOTE-OWELM, ISMOTE-OWELM, and LSTM-SAE methods have resulted in a moderately closer accuracy of 0.918, 0.920, and 0.930. But the proposed OLSTM-SAE technique has accomplished superior outcomes with an accuracy of 0.932. On exploring the results interms of F-score, it is evident that LDT/UDT, SVM, and PCPM approaches have attained a minimum F-score of 0.670, 0.731, and 0.831. Similarly, the WELM and OWELM models have attained slightly enhanced F-score of 0.872 and 0.894. In line with, the SMOTE-OWELM, ISMOTE-OWELM, and LSTM-SAE techniques have resulted in a moderately closer F-score of 0.917, 0.918, and 0.931. But the proposed OLSTM-SAE manner has accomplished higher outcome with an F-score of 0.943.

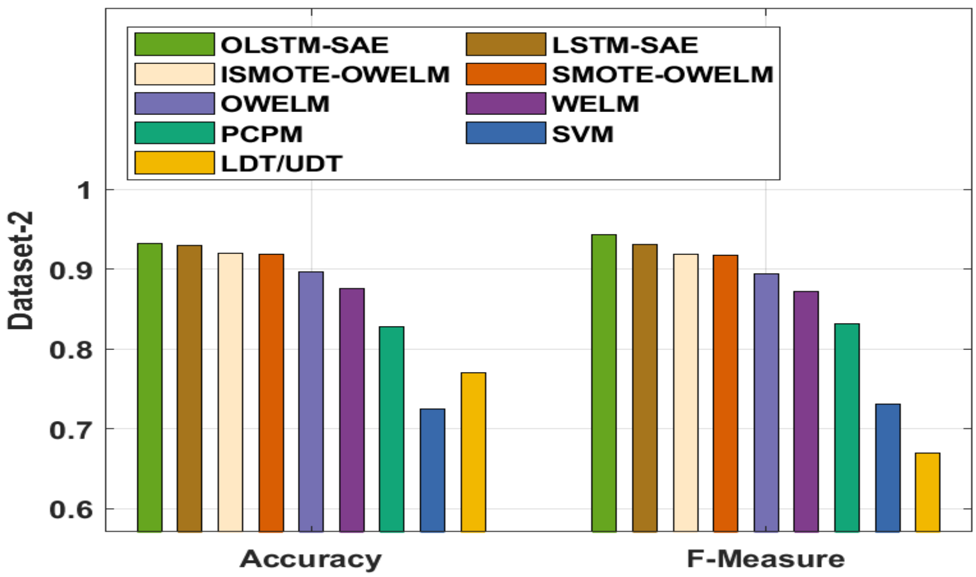

Fig. 3. Comparison study of OLSTM-SAE technique on dataset-2

DOI: 10.5281/zenodo.5502106
Received: May 01, 2021 Accepted: August 29, 2021

Fig. 4 considers the results analysis of the OLSTM-SAE algorithm accuracy, it can be clear that LDT/UDT, SVM, and PCPM techniques have obtained a lower accuracy of 0.580, 0.679, and 0.818. Followed by, the WELM and OWELM methodologies WELM, ISMOTE-OWELM, and LSTM-SAE manners have resulted in a moderately closer accuracy of 0.899, 0.909, and 0.914. However, the presented OLSTM-SAE system has accomplished maximal outcomes with an accuracy of 0.927. On considering the outcomes with respect to F-score, it can be demonstrated that LDT/UDT, SVM, and PCPM techniques have achieved a reduced F-score of 0.592, 0.682, and 0.808. At the same time, the WELM and OWELM models have attained slightly enhanced F-score of 0.852 and 0.879. In line with, the SMOTE-OWELM, ISMOTE-OWELM, and LSTM-SAE manners have resulted in a moderately closer F-score of 0.899, 0.908, and 0.913. But the proposed OLSTM-SAE algorithm has accomplished maximum result with an F-score of 0.931.

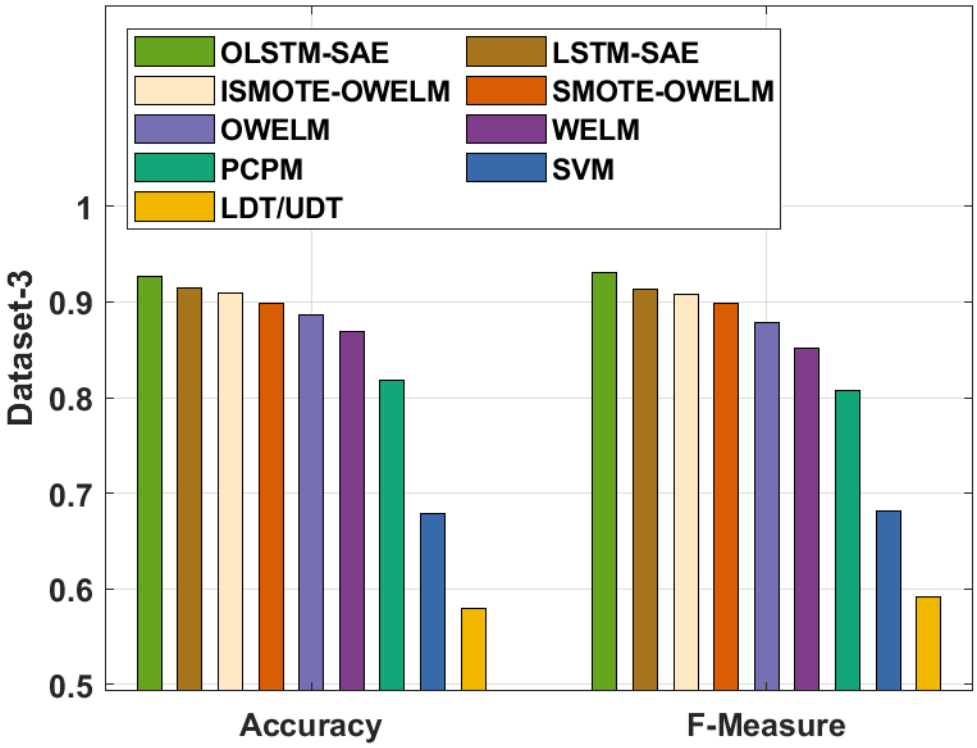

Fig. 4. Comparison study of OLSTM-SAE technique on dataset-3

5. Conclusion

This paper has presented an effective OLSTM-SAE technique for CCP in telecom industry. The OLSTM-SAE technique encompasses three subprocesses namely preprocessing, classification, and parameter optimization. The OLSTM-SAE technique classifies the preprocessed data into churn and non-churn customers. In addition, the GWO technique is used to adjust the variables involved in the LSTM-SAE model. For examining the enhanced performance of the OLSTM-SAE technique, an extensive simulation analysis takes place and the outcomes are inspected with respect to various measures. The experimental results highlighted the betterment of the OLSTM-SAE technique interms of different evaluation parameters. In future, the performance of the OLSTM-SAE technique can be boosted using feature selection processes.

References

[1] Ahmed, A.A. and Maheswari, D., 2017. Churn prediction on huge telecom data using hybrid firefly based classification. Egyptian Informatics Journal, 18(3), pp.215-220.

[2] Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J. and Anwar, S., 2019. Customer churn prediction in telecommunication industry using data certainty. Journal of Business Research, 94, pp.290-301.

[3] Idris, A., Iftikhar, A. and ur Rehman, Z., 2019. Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. Cluster Computing, 22(3), pp.7241-7255.

DOI: 10.5281/zenodo.5502106
Received: May 01, 2021 Accepted: August 29, 2021

[4] Alboukaey, N., Joukhadar, A. and Ghneim, N., 2020. Dynamic behavior based churn prediction in mobile telecom. Expert Systems with Applications, 162, p.113779.

[5] Azeem, M., Usman, M. and Fong, A.C.M., 2017. A churn prediction model for prepaid customers in telecom using fuzzy classifiers. telecommunication Systems, 66(4), pp.603-614.

[6] Khan, Y., Shafiq, S., Naeem, A., Hussain, S., Ahmed, S. and Safwan, N., 2019. Customers churn prediction using artificial neural networks (ANN) in telecom industry. Editorial Preface From the Desk of Managing Editor, 10(9).

[7] Mishra, A. and Reddy, U.S., 2017, December. A novel approach for churn prediction using deep learning. In 2017 IEEE international conference on computational intelligence and computing research (ICCIC) (pp. 1-4). IEEE.

[8] Amin, A., Shah, B., Khattak, A.M., Moreira, F.J.L., Ali, G., Rocha, Á. and Anwar, S., 2019. Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. International Journal of Information Management, 46, pp.304-319.

[9] Mishra, A. and Reddy, U.S., 2017, November. A comparative study of customer churn prediction in telecom industry using ensemble based classifiers. In 2017 International Conference on Inventive Computing and Informatics (ICICI) (pp. 721-725). IEEE.

[10] Zhao, L., Gao, Q., Dong, X., Dong, A. and Dong, X., 2017. K-local maximum margin feature extraction algorithm for churn prediction in telecom. Cluster Computing, 20(2), pp.1401-1409.

[11] Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A. and Huang, K., 2017. Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing, 237, pp.242-254.

[12] Prashanth, R., Deepak, K. and Meher, A.K., 2017, July. High accuracy predictive modelling for customer churn prediction in telecom industry. In International Conference on Machine Learning and Data Mining in Pattern Recognition (pp. 391-402). Springer, Cham.

[13] Spanoudes, P. and Nguyen, T., 2017. Deep learning in customer churn prediction: unsupervised feature learning on abstract company independent feature vectors. arXiv preprint arXiv:1703.03869.

[14] Cenggoro, T.W., Wirastari, R.A., Rudianto, E., Mohadi, M.I., Ratj, D. and Pardamean, B., 2021. Deep Learning as a Vector Embedding Model for Customer Churn. Procedia Computer Science, 179, pp.624-631.

[15] Pustokhina, I.V., Pustokhin, D.A., Aswathy, R.H., Jayasankar, T., Jeyalakshmi, C., Díaz, V.G. and Shankar, K., 2021. Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms. Information Processing & Management, 58(6), p.102706.

[16] Domingos, E., Ojeme, B. and Daramola, O., 2021. Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector. Computation, 9(3), p.34.

[17] Tariq, M.U., Babar, M., Poulin, M. and Khattak, A.S., 2021. Distributed model for customer churn prediction using convolutional neural network. Journal of Modelling in Management.

[18] De Caigny, A., Coussement, K., De Bock, K.W. and Lessmann, S., 2020. Incorporating textual information in customer churn prediction models based on a convolutional neural network. International Journal of Forecasting, 36(4), pp.1563-1578.

[19] Kozak, J., Kania, K., Juszczuk, P. and Mitręga, M., 2021. Swarm intelligence goal-oriented approach to data-driven innovation in customer churn management. International Journal of Information Management, p.102357.

[20] Jin, Z., Yang, Y. and Liu, Y., 2020. Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 32(13), pp.9713-9729.

[21] Li, Y. and Cao, H., 2018. Prediction for tourism flow based on LSTM neural network. Procedia Computer Science, 129, pp.277-283.

[22] Mirjalili, S., Mirjalili, S.M. and Lewis, A., 2014. Grey wolf optimizer. Advances in engineering software, 69, pp.46-61.

[23] Zareie, A., Sheikhahmadi, A. and Jalili, M., 2020. Identification of influential users in social network using gray wolf optimization algorithm. Expert Systems with Applications, 142, p.112971.

[24] Dadashzadeh, S., Aghaie, M. and Zolfaghari, A., 2021. Implementation of Gray Wolf Optimization algorithm to recycled gas centrifuge cascades. Progress in Nuclear Energy, 137, p.103769.

DOI: 10.5281/zenodo.5502106
Received: May 01, 2021 Accepted: August 29, 2021