

Deep Learning-based Churn Prediction of Telecom Subscribers

Nasebah Almfadi¹, Ali Mustafa Qamar^{1,3}, Rehan Ullah Khan^{2,4}, Mohamed Tahar Ben Othman^{1,3}

¹ Department of Computer Science, College of Computer, Qassim University, Al-Mulaida, 51431, Saudi Arabia

² Department of Information Technology, College of Computer, Qassim University, Al-Mulaida, 51431, Saudi Arabia

³ BIND Research Group, College of Computer, Qassim University, Al-Mulaida, 51431, Saudi Arabia

⁴ Intelligent Analytics Group (IAG), College of Computer, Qassim University, Al-Mulaida, 51431, Saudi Arabia

Abstract

During the last decade, mobile communication has become the most sought-after mode of communication. However, the telecommunication market has reached a point of saturation and it is, in general very difficult for a company to bring a new customer. On the other hand, mobile number portability has become very easy. This has resulted in an ever-increased concept of churn, whereby a customer moves from one service provider to another. It is very important for telecommunication companies to predict which of the customers are most susceptible to move to their competitors. One can predict this move based on Call Detail Records (CDR). CDRs have features such as last call date, duration of a call, and data usage. Deep Learning is a relatively recent concept and has found many applications in pattern recognition, image processing, and other related domains. As such, in this research, we apply deep learning for churn prediction. For evaluation and analysis, we use the Mobile Telephony Churn Prediction Dataset. The dataset contains the data for around 100k individuals, out of which 86k individuals are non-churners whereas, 14k are the churned ones. From the evaluation of the proposed deep model, we achieve a 96% accurate model for the detection of churn and non-churn instances, thus, showing the efficacy of the approach for practical applications.

Keywords: Deep learning, Machine Learning, Data mining, Churn prediction

I. INTRODUCTION

With ever-increasing competition in the telecommunication industry, predicting customer churn for various service providers has become a demanding problem. The moving away of a registered user from a service provider is known as customer churn. It is a widely known fact that retaining an existing customer is cheaper than acquiring a new one. Moreover, long-term customers help generate higher profits as compared to new customers. Therefore, it is important for a company to predict customer churn in advance. It helps to retain the existing customers by taking active steps to curtail churn [1]. With the growth of the data, machine learning has emerged

as a significant paradigm for mining the data because of its ability to extract patterns and then learning these patterns for the prediction of new instances [2]. The deep learning has received much attention during the last decade. A lot of deep learning algorithms have been developed to enhance machine learning algorithms to find multiple levels of representations. A deep learning algorithm is a special type of Neural Network which models the various levels of representation, thus higher-level features are used to represent more abstract aspects of the data [3]. In this article, deep learning techniques are applied to a telecommunication dataset obtained from “customers-dna.com” to predict customer churn for telecommunication service providers. The customer churn analysis can help an organization in making business decisions and expand their services. The deep learning model can be applied to various areas and the insights obtained from deep learning-based churn prediction can contribute to problems in many other similar domains.

II. RELATED WORK

Many machine learning algorithms have been used to solve the customer churn prediction problem. Yuan et al. [1] built a customer churn prediction model based on the Juzu Entertainment app. 100,000 user samples were analyzed and 75 percent of churners were identified by the model. The results showed that the proposed customer churn method achieved good results. Gordini and Veglio [2] developed a churn prediction model for an Italian B2B e-commerce organization. The dataset contains data of 80,000 customers. The forecasting nature of Support Vector Machines (SVM) based on the Area under the ROC curve (AUC) parameter selection technique was explored. The proposed approach was compared to SVM, Artificial Neural Networks (ANNs) and Logistic Regression algorithms. The results showed that SVM was the most successful algorithm to predict churn. Coussemant and Bock [4] worked on a real-life online gambling dataset provided by BWIN Interactive Entertainment. The research confirmed that predicting customer churn is crucial in customer relationship management (CRM) portfolio of a gambling company. The study showed that the ensemble algorithms should be preferred

over standard algorithms to solve the problem of customer churn. Zhu et al. [5] demonstrate a detailed comparison of the state-of-the-art approaches for class imbalance on real-world churn prediction datasets. Based on the experimental results, the research provides some guidance to choose the best approaches to overcome the class imbalance problem. Ahmed and Linen [6] carried out a detailed review of churn prediction techniques developed for the telecommunication industry. Most of the current churn prediction algorithms use machine learning and meta-heuristic approaches. Some researchers used algorithms such as SVM and ANN, others tried to improve data samples through effective pre-processing. It has been found that the most accurate churn prediction is accomplished when applying hybrid methods rather than individual algorithms. Hybrid algorithms, either the hybrid machine learning ones or hybrid meta-heuristics are better suited to predict churn successfully. Esteves and Mendes-Moreira [7] built various models using k-Nearest Neighbor (kNN), Naïve Bayes, Random Forest, in order to detect the customers who are closer to abandoning their telecom service providers using a hybrid sampling method based on the Synthetic Minority Over-Sampling Technique (SMOTE). The models were evaluated based on AUC, sensitivity, and specificity. Random Forest achieves the best results among all the methods. Qureshi et al. [10] used a number of data mining algorithms for churn prediction such as regression analysis, decision trees, and ANN. The authors also used the dataset from “customers-dna.com” by finding patterns from the historical data to distinguish potential churners.

In recent years, there has been considerable interest in the prediction of churn using deep learning. Deep learning is an evolving field within the machine learning research community. Deep learning algorithms help in the discovery of multiple levels of representation [3]. The authors in [11] summarize historical surveys and relevant works from state-of-the-art. Recently, researchers have started to study deep learning models to improve the classification effectiveness in the area of telecom churn prediction. Prashanth et al. [8] studied statistical and data mining techniques for predicting churn. The proposed approach uses linear and logistic regression as well as non-linear methods of Random Forest and Deep Learning architecture including Deep Neural Network, Deep Belief Networks (DBN) and Recurrent Neural Networks (RNN) for churn prediction. They were the pioneers for predicting churn using both conventional machine learning methods as well as deep learning algorithms. They found that the non-linear models perform better as compared to other approaches and predictive models could be used in the telecom industry to make better and informed decisions and customer management. Another study in this field was conducted by Umayaparvathi and Iyakutti [9] who used two telecom datasets and applied various deep neural network architectures, such as Feedforward Neural Network (FNN), Convolutional Neural Network (CNN). Moreover, the authors also performed classification using traditional machine learning algorithms. The

experimental results illustrated that deep learning models performed as good as the conventional machine learning algorithms.

III. METHODOLOGY

Customer churn has become a critical issue for decision-makers. Machine learning tools are among the most widely used ones in studies related to customer churn prediction. Few researchers have addressed the problem from a deep learning perspective. However, there is still a need for advanced deep learning tools for solving the customer churn problem. The present paper aims to develop a deep learning model based on Convolutional Neural Networks (CNN) architecture to effectively detect customer churn in a telecommunication company. To the best of our knowledge, this will be the first study that applies deep learning on this dataset. Fig. 1 presents the flowchart of the work.

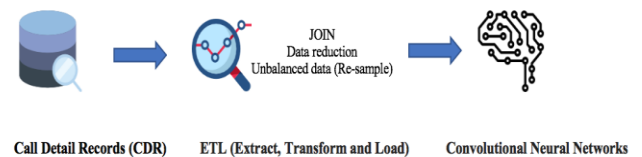


Fig. 1. Flowchart of the proposed approach

III.I Dataset

The dataset used in this research is obtained from a Telecom operator, containing three months of customer call detail records. The dataset contains two classes, represented as the active and churn instances. Thus, the dataset represents a bi-class problem. The dataset contains 48 attributes. The main attributes are traffic type (outgoing/ incoming for voice, SMS, and data), traffic destination (on-net, off-net), rate plan, customer loyalty, and traffic behavior.

III.II Imbalanced Data

The accuracy-based evaluation does not work properly with imbalanced data. The key problem with imbalanced data is that even though the overall accuracy is high, yet many examples may have been wrongly classified as belonging to the majority class. Our dataset is an imbalanced dataset as shown in Fig. 2. In this paper, the Resample method is used, which is one of the several techniques to under-sample a dataset used in a classification problem. In this technique, the samples are removed from the majority class to balance the dataset.

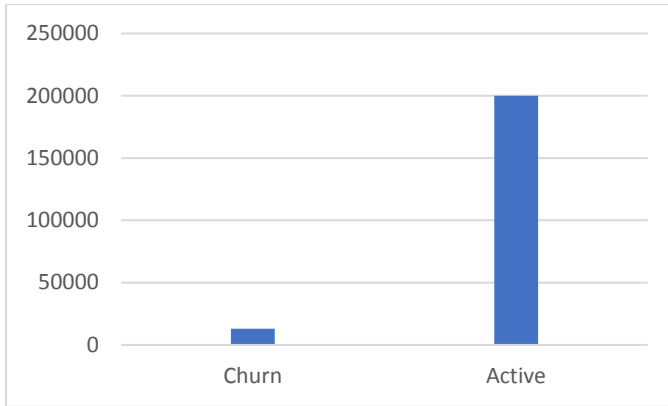


Fig. 2. Histogram of the *Status* attribute. The y-axis shows the frequency

During the pre-processing stage, the first step was to apply the under-sampling method that works by reducing the majority class samples i.e. the *Active* class. This reduction is done in a random manner, in which case it is called random under-sampling. The Resample filter available in *WEKA* software is employed to under-sample the majority class. The parameter *sampleSizePercent* is set as 20, which represents the size of the output dataset as a percentage of the input dataset. Fig. 3 demonstrates the dataset after reduction.

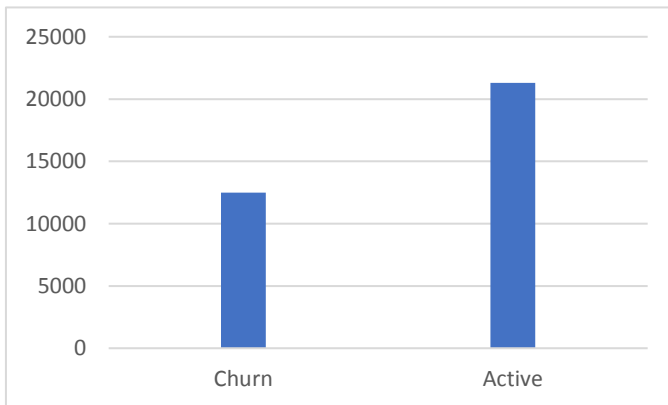


Fig. 3. Histogram of the *Status* attribute after under-sampling *Active* class. The y-axis shows the frequency.

III.III Data Pre-processing

In this sub-section, the main pre-processing steps followed in this work are discussed:

1. Feature Selection

A closer look at the data indicates that not all of the attributes are effective for building the model. The *Auto Model* available in RapidMiner was used to select effective attributes.

2. Data Transformation

Many models cannot operate on labeled data directly. They require transforming categorical features into numerical ones. In this case, one hot encoding was employed.

3. Handling Missing Values

The missing values in a dataset must be properly compensated. This problem was resolved by calculating the average of the individual attributes followed by the replacement of the missing values for each attribute.

4. Normalization

After executing the model, the results proved that normalization was an important step to deal with overfitting issues.

III.IV Proposed Deep Architecture

Hyper-parameters are the variables that help decide how a deep neural network is trained. The values of various hyper-parameters were selected based on the results of testing different models. Table 1 indicates the hyper-parameter settings for the proposed CNN model.

The CNN model is composed of two convolutional layers, which contains 512 filters of kernel size 1. In order to break linearity, the Rectified Linear Units (ReLU) function is used to compute the activation for the convolutional layer. Similarly, the feature maps of the convolutional layer will be the input of the pooling layer, and they will be pooled to reduce the feature dimensions and to accelerate the computation. Max-Pooling (with pool size 1) is used. Furthermore, Flatten and dropout having a rate of 0.5 was applied between each Fully Connected (FC) layer to avoid over-fitting. After the computation of several convolutional and pooling layers, the output of the last layer is the input of an FC layer, which will help to perform the classification. The subsequent processing of data is performed by FC layers with the ReLU activation function. Since this is a binary classification problem, the output of the FC layer will be computed by the activation function of the sigmoid. ADAM is used to adapt the learning rate, whereas binary cross-entropy is used as the objective function. The number of epochs is 30 which represents the number of training iterations over the dataset. Once the parameter selection was completed, a predictive model was trained and evaluated using the first two months of data from the customer's CDR.

Table 1. Hyper-parameter settings for the proposed model

Parameters	Values
Number of convolutional layers	2
Number of dense layers (FC)	1
Epochs	30
Number of filters	[32, 64]
Dropout rate	0.5
Optimizer	Adam
Learning rate	10^{-4}
Hidden neurons	[64]
Loss function	binary cross-entropy
Activation function (CNN)	ReLU
Activation function (FC)	ReLU
Activation function (output)	Sigmoid
Total number of parameters	178,433

The architecture of CNN is shown in Fig. 4. The model consists of an input layer, two convolutional (Conv.) layers, two pooling layers, one fully connected (FC) layer, and an output layer.

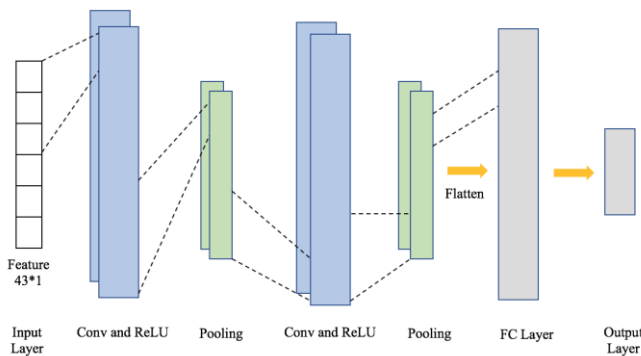


Fig. 4. The architecture of the proposed CNN Model

IV. EXPERIMENTS AND RESULTS

This section shows the evaluation results and discusses the performance of churn prediction classifiers.

After constructing the model, the next step is to fit the CNN model to the dataset. We used three months of data containing 49983 instances. The data for the first two months was employed for training and validation (with 80:20 split), resulting in 27008 samples for training and 6753 instances for validation. Similarly, the third month was used for testing containing 16222 samples.

In order to measure the performance, various metrics such as accuracy, precision, recall, F-measure, Receiver Operating Characteristic (ROC), and AUC precision-recall curve (PRC) were considered for evaluation. Table 2 summarizes the results obtained on the test data. Similarly, Table 3 shows the confusion matrix for the model.

Table 2. Results on the test data

Metric	Proposed Churn Model
Accuracy	0.965
Precision	0.917
Recall	0.998
F1-score	0.956
ROC AUC	0.995
AUC Precision-Recall Curve (PRC)	0.991
Cohens kappa	0.927

Table 3. Confusion matrix

		Predicted Class	
		Active Customers	Churners
Actual Class	Active Customers	9434	557
	Churners	8	6223

The results obtained with the proposed model are much better as compared to the ones obtained by Qureshi et al. [10] on the same dataset. The proposed strategy results in an improvement of at least 10% in the recall, which can be considered as a significant increase. The training and validation loss is presented in Fig. 5, whereas Fig. 6 shows the accuracy obtained on the training and validation sets. The figures show the reduction of loss and the increase in the accuracy for training and validation sets, respectively as the number of epochs is increased. One can easily observe from the loss curve in Fig. 5 that the results on the validation set are better as compared to

the training set. On the other hand, the difference between the training and the validation accuracies decrease as the number of epochs is increased as shown in Fig. 6.

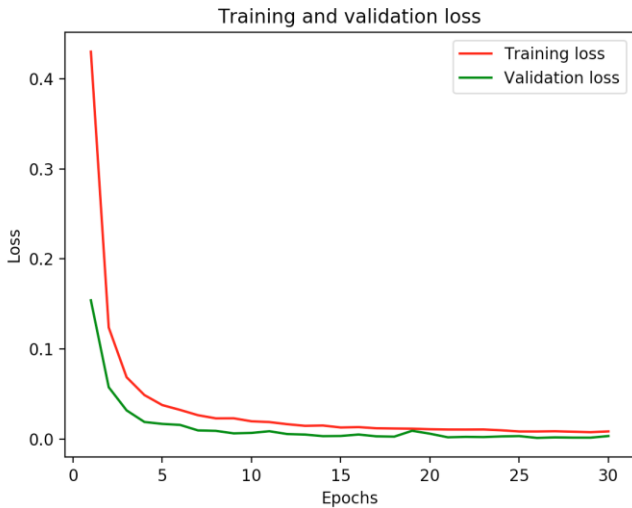


Fig. 5. Training and validation loss curve for the best accurate model where training loss is 0.0085 and validation loss is 0.0046

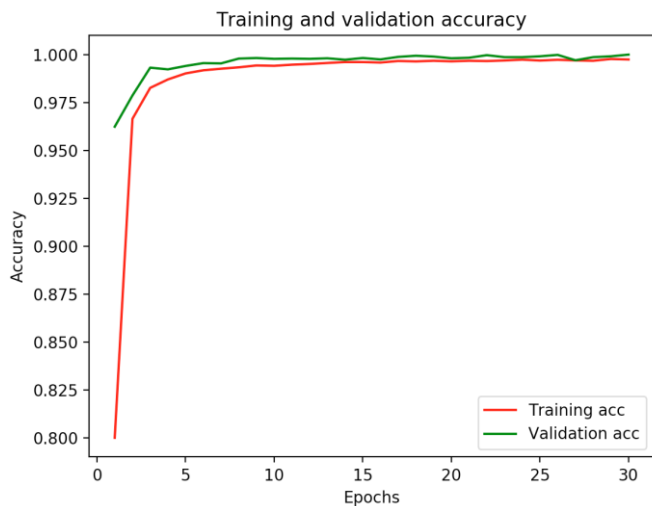


Fig. 6. Training and validation accuracy curve for the best accurate model where training accuracy is 0.9973 and validation accuracy is 0.9979

In order to further evaluate the prediction of the proposed approach, we illustrate the Precision-Recall Curve (PRC) in Fig. 7. Furthermore, the ROC curve obtained on the test set while using the ReLU nonlinearity function is shown in Fig. 8.

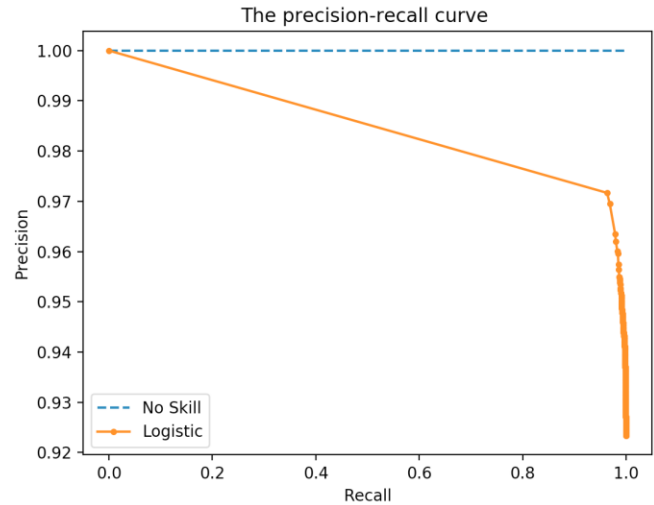


Fig. 7. The precision-recall curve for the test set

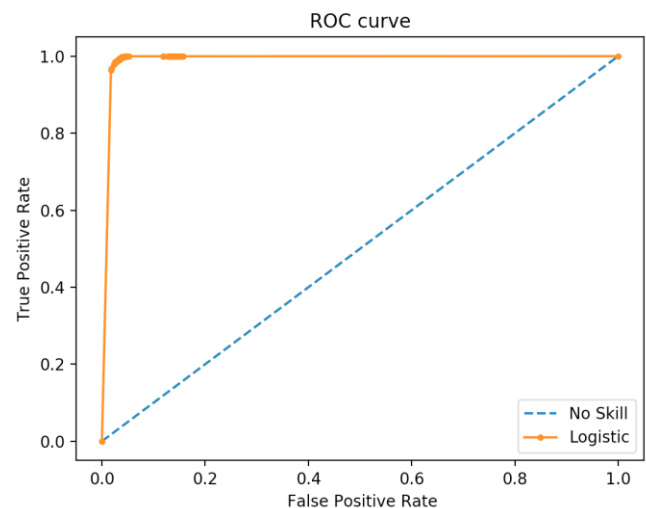


Fig. 8. ROC curve for the test set

The results reported here allow us to draw the conclusion that 1D CNN was not only able to work with the numeric data but also performed classification with very high accuracy. Moreover, the use of ReLU as an activation function further increases the classification accuracy.

Further analysis was performed in order to compare the prediction performance between the proposed model and other classifiers using different evaluation metrics. We evaluated four models for the churn prediction task as highlighted in Table 4. By analyzing the obtained results, the most remarkable results are with decision tree classifier J48. The experimental results show that the proposed deep learning-based churn prediction model performs as good as the most common classification algorithms.

Table 4. Comparison with various state-of-the-art classification techniques

Classification Technique	Accuracy	Precision	Recall	F1-score	PRC	ROC
ANN	0.997	0.997	0.997	0.997	1.0	1.0
kNN	0.964	0.966	0.964	0.965	0.952	0.967
Naïve Bayes	0.971	0.972	0.972	0.972	0.990	0.994
J48	1.0	1.0	1.0	1.0	1.0	1.0

V. CONCLUSION AND FUTURE WORK

The problem of churn has affected a lot of sectors, telecommunication is one of them. Getting a new customer is much more expensive as compared to retaining the existing ones. In this research work, a convolutional neural network was developed to effectively address the problem of churn in a telecommunication company. This research presented a new paradigm, providing a 1D CNN model for churn prediction that predicted customers with a high propensity to churn. The deep learning model to detect churn and non-churn customers was found to be 96% accurate. It is strongly believed that this model has the potential for making better decisions for churn management in the telecom industry. This work is an attempt to use only one dataset for predictive modeling. In future, one can build similar models with other datasets. The proposed approach could be used in areas other than telecommunication.

Acknowledgments: The authors gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research, on the material support for this research under the Project number: 3502-coc-2018-1-14-S during the academic year 1439 AH / 2018 AD.

REFERENCES

1. Yuan S, Bai S, Song M, Zhou Z. Customer Churn Prediction in the Online New Media Platform: A Case Study on Juzi Entertainment. 2017 Int Conf Platf Technol Serv PlatCon 2017 - Proc. 2017; (8).
2. Gordini N, Veglio V. Customers churn prediction and

- marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Ind Mark Manag.* 2017, 62:100–7
3. Bengio Y. Deep learning of representations: Looking forward. In: *International Conference on Statistical Language and Speech Processing, Lect Notes Computer Science*, 7978, 1–37, 2013
4. Coussement K, Bock KW De. Customer churn prediction in the online gambling industry : The beneficial effect of ensemble learning. *J Bus Res.* 2013;66(9):1629–36.
5. Zhu B, Baesens B, vanden Broucke SKLM. An empirical comparison of techniques for the class imbalance problem in churn prediction. *Inf Sci (Ny).* 2017;408:84–99.
6. Ahmed A, Linen DM. A review and analysis of churn prediction methods for customer retention in telecom industries. In: *4th International Conference on Advanced Computing and Communication Systems, ICACCS*, 2017.
7. Esteves G, Mendes-Moreira J. Churn prediction in the telecom business. In: *11th International Conference on Digital Information Management (ICDIM)*, 2016. 254–9.
8. Prashanth R, Deepak K, Meher AK. High Accuracy Predictive Modelling for Customer Churn Prediction in Telecom Industry. In: *International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM)*, 2017. 1–13.
9. Umayaparvathi V, Iyakutti K. Automated Feature Selection and Churn Prediction using Deep Learning Models. *Int Res J Eng Technol*, 4(3), 1846–54, 2017
10. Qureshi SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: *8th International Conference on Digital Information Management (ICDIM)*, 2013. 131–6.
11. Schmidhuber J. Deep Learning in Neural Networks: An Overview. *Expert Syst Appl.* 2017;72:221–30.