DOI: 10.24996/ijs.2022.63.6.32

Iraqi Journal of Science

ISSN: 0067-2904

Telecom Churn Prediction based on Deep Learning Approach

Israa N. Mahmood*, Hasanen S. Abdullah
Department of Computer Science, University of Technology, Baghdad, Iraq

Received: 23/3/2021 Accepted: 22/8/2021 Published: 30/6/2022

Abstract
The transition of customers from one telecom operator to another has a direct impact on the company's growth and revenue. Traditional classification algorithms fail to predict churn effectively. This research introduces a deep learning model for predicting customers planning to leave to another operator. The model works on a high-dimensional large-scale data set. The performance of the model was measured against other classification algorithms, such as Gaussian NB, Random Forrest, and Decision Tree in predicting churn. The evaluation was performed based on accuracy, precision, recall, F-measure, Area Under Curve (AUC), and Receiver Operating Characteristic (ROC) Curve. The proposed deep learning model performs better than other prediction models and achieves a high accuracy rate of 91%. Furthermore, it was noticed that the deep learning model outperforms a small size Neural Network for the customer churn prediction.

Keywords: churn prediction, deep learning, classification algorithms

التنبؤ بتضاؤل أعداد المشتركين في مجال الأتصالات بأستخدام تقنية التعلم العميق

اسراء نافع محمود* ، حسنين سمير عبد الله*
قسم علوم الحاسوب، الجامعة التكنولوجية،بغداد، العراق

الخلاصة
إن انتقال العملاء من مشغل اتصالات إلى آخر له تأثير مباشر على نمو الشركة وإيراداتها. تفشل خوارزميات التصنيف التقليدية في التنبؤ بفاعلية. يقدم هذا البحث نموذجًا تعليميًا عميقًا لتخمين العملاء الذين يخططون للمغادرة إلى مشغل آخر. يعمل النموذج على مجموعة بيانات كبيرة الحجم وعالية الأبعاد. تم قياس اداء النموذج مقابل خوارزميات التصنيف الأخرى، وهي Gaussian NB و Random Forrest و Decision Tree. تم إجراء التقييم بناءً على عدة مقاييس مثل الدقة، الضبط، الاسترجاع، القياس (f measure)، منحنى خصائص تشغيل المستقبل (ROC) و المنطقة تحت المنحى (AUC). اظهرت النتائج ان التعلم العميق يحقق معدل دقة عالياً (٩١٪) مقارنة بخورزاميات التصنيف الاخرى. علاوة على ذلك، لوحظ ان التعلم العميق يتفوق علي الشبكة العصبية صغيرة الحجم في التنبؤ بتناقص المشتركين.

1. Introduction

is the event when a subscriber leaves one telecom operator to a rival operator in the market. The number of churn users in a corporation is considered one of the most significant factors

*Email: cs.19.78@grad.uotechnology.edu.iq

that affect the corporation's financial status. Customer Churn analysis is extremely important for project managers to build a strategy that enhances the company's growth and increases its sales [1].
Recent studies show that retaining customers is more profitable than gaining new customers [2, 3]. Over the last few years, different industries mainly in Banking, telecommunication, airlines, social network, as well as companies that offer services based on subscription, such as online entrainment like Netflix have focused on predicting churn subscribers [4, 5, 6].
Building a highly accurate model for churn prediction is a fundamental part of telecom companies planning process and strategic decision-making. This has led the competitive companies to develop predictive tools that help in classification and predictive modeling.
Recently, there has been an increase in the volume and sources of information that is handled to elicit knowledge and insights [7]. This rise has forced the researchers and developers to employ new techniques like deep learning instead of older techniques that are less accurate with Big data, such as Support Vector Machines (SVM), Naïve Bayes, logistic regression, and
Of what?
decision tree. This study will focus on extracting knowledge from a huge telecom dataset using a deep learning model with Keras library.
The aim of this research is to develop a high accuracy deep learning model that predicts churn subscribers without causing overfitting. A high dimensional big data dataset will be used to train and test the model.
This paper is organized as follows; different methods of churn prediction in telecommunication are presented in the literature review in Section 2. Section 3 explains how the dataset is constructed and the proposed system of prediction. Section 4 presents the finding of the proposed system and how the deep learning model was able to achieve high prediction without overfitting compared to the old techniques. Section 5 summarizes the conclusion and the direction of future work.

2. Literature Review

In the past decade, several studies were conducted in the field of churn prediction by researchers in academic and industry fields. Different approaches were implemented to predict churn in many sectors. Various examples from these studies regarding churn prediction in the telecommunication sector are discussed in this section.
Kumar et al. [8] proposed a model to predict churn that uses an Artificial Bee Colony (ABC) algorithm to train the Artificial Neural Network (ANN). The weights of the neural network are evolved using the ABC algorithm. The model achieves the minimum error rate of 0.11 compared to other algorithms used for churn prediction.
Raza et al. [9] suggested a churn prediction model using the Random Forrest algorithm. The proposed model provides a group-based retention offer to churn customers by dividing them into multiple groups using the k-mean clustering algorithm. The suggested model was able to correctly classify instances with 88.63% accuracy.
Naeem et al. [10] implemented an ANN that predicts churn customers for Pakistan telecom operators that achieves a 79% accuracy rate. A backpropagation algorithm was used to train the model and an analysis of the importance of each variable was provided to estimate the influence of each aspect. Moreover, Sivasankar et al. [11] introduced a methodology that recognizes the important features in predicting churn and uses these features in ensemble-classification algorithms like Boosting, Random Subspace, and Bagging. The suggested model that combines feature selection with ensemble classification attained a 95.13% accuracy rate. Additionally, Idris et al. [12] developed a model that uses the Particle Swarm Optimization based undersampling technique to handle the problem of imbalanced class distribution of Telecom dataset and GP-AdaBoost Algorithm to perform binary classification on the balanced dataset. The proposed system yields 0.91 Area Under the Curve (AUC).

Lu et al. [13] used an AdaBoost Algorithm to divide the customers into 2 clusters (Churn, Non-Churn) and used a logistic regression algorithm as a churn prediction model for each cluster. The experimental results show that a good separation of churn data was provided using boosting. Dahiya et al. [14] also discussed different prediction models and compared the results of prediction models like logistic regression and J-48 Decision tree when using various size datasets.

Amin et al. [15] introduced a novel customer churn prediction approach based on the concept that different zones in the dataset have different accuracy levels. The correlation can be found between the accuracy of the classifier and prediction certainty using the distance factor. The approach divides the dataset into two categories (data with low certainty, data with high certainty) for predicting churn and non-churn. Additionally, Ammar et al. [16] developed a heterogeneous ensemble classifier that employs several classification algorithms to perform the first-level prediction. A second-level prediction utilizing a heuristic-based combiner was used for prediction with discrepancies to provide a final prediction. The proposed model saves cost by almost 50% by using the techniques of customer uplift modeling. Panigrahi et al. [17] also implemented a neural network for predicting churn and recognized features that have a high impact on predicting churn.

To the best of our knowledge in the literature review, we were not able to locate research that works on the development and implementation of a Deep Learning prediction model for high-dimensional large-scale telecom dataset. This paper aims to construct a deep learning model that achieves a high prediction rate without affecting the generalization of the model.

3. The Proposed Model of Churn Prediction

Churn prediction is a binary classification issue as it encompasses two classes: churn and non-churn. The churn prediction model focuses on providing insights into the customer's future within the company.

An ultimate understanding of the past activity of the customers is required to identify the future prospect.

3.1 Data Set

In this paper, we choose to use a real data set from the telecom industry. The telecom operator provided us with Call Detailed Record (CDR) for more than 2 years. CDR is a record generated by the telecom billing system that includes details about subscriber activities. Since the CDRs cannot be applied directly to the churn prediction model. The data was first prepared and preprocessed to produce a data set that consisted of more than 600 attributes and one target attribute "Churn" for more than 2.5 million subscribers. As part of data preparation, the unique features were removed from the dataset and the categorical features were changed into numerical. In data preprocessing, the missing values were filled, and noise and outliers removed. The data was partitioned into training and testing datasets with a test size equal to 20%.

3.2 Proposed Model Design

Deep Learning is a machine learning method that mimics the behavior of the human brain in detecting objects, processing data, translating language, making decisions, and many other different fields. One of the most significant characteristics of deep learning that differentiate it from traditional machine learning algorithms is that there is no need to extract features using classical feature selection algorithms.

Instead, the features are learned automatically with no need to apply a separate algorithm, and new complex features are extracted from the simple features to make a better decision. The proposed system is shown in Figure 1.

We build a Deep Neural Network that consists of an input layer for 667 features, 11 hidden layers with different numbers of neurons at each layer, and 1 output layer. The architecture of the model is shown in Table 1.

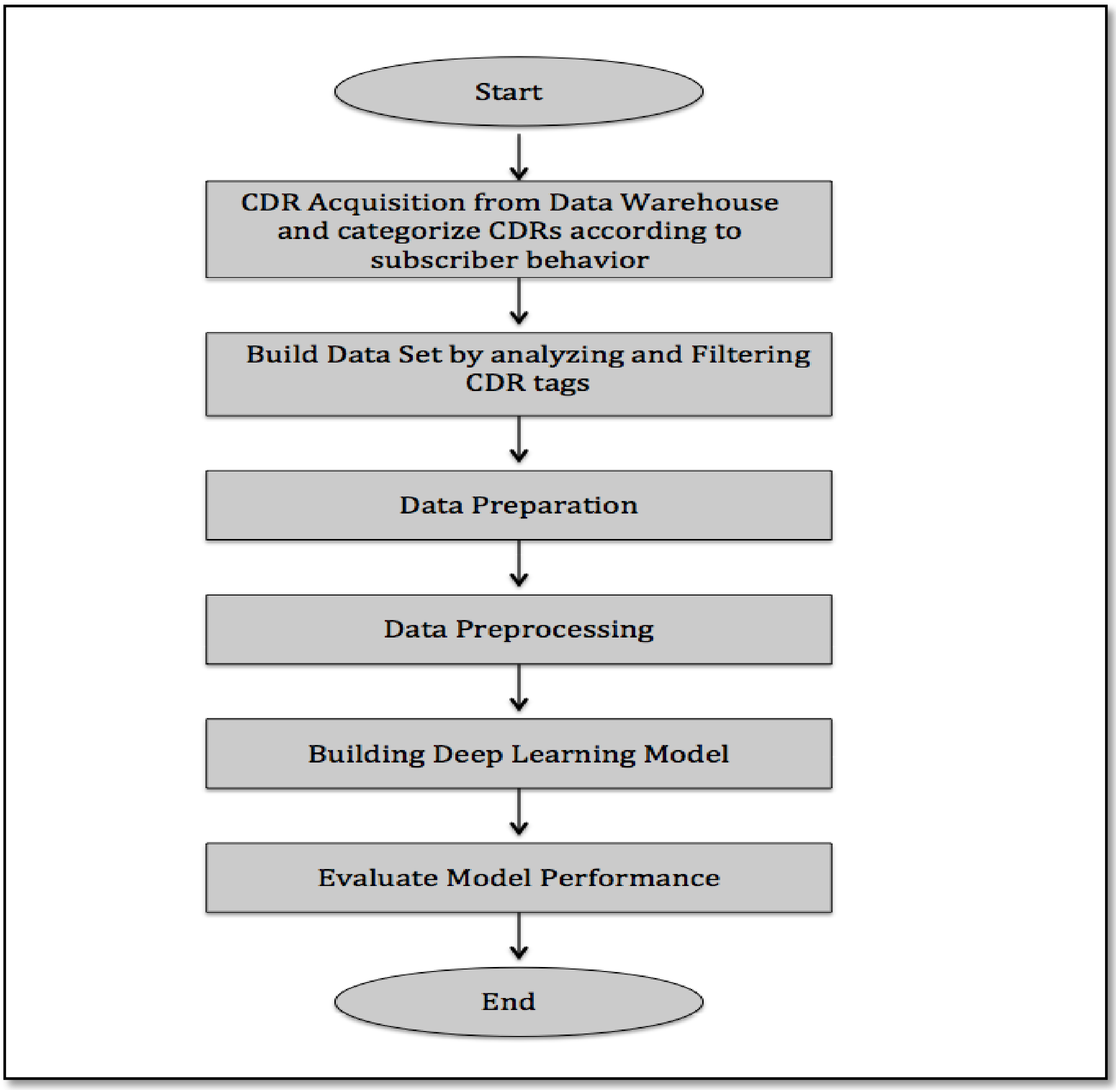

Figure 1-Proposed System Framework Design

Table 1- Deep Learning Model Architecture

Layer	Number of Neurons	Activation Function
Hidden Layer 1	128	Relu
Hidden Layer 2	64	Relu
Hidden Layer 3	64	Relu
Hidden Layer 4	32	Relu
Hidden Layer 5	32	Relu
Hidden Layer 6	32	Relu
Hidden Layer 7	32	Relu
Hidden Layer 8	16	Relu
Hidden Layer 9	16	Relu
Hidden Layer 10	16	Relu
Hidden Layer 11	16	Relu
Output Layer	1	Sigmoid

4. Proposed Model Results Analysis and Discussion

A detailed description of the measures that can be used to assess the model performance is included in Section 4.1. The results of the proposed system against other traditional prediction models are discussed in Section 4.2.

4.1 Evaluation Measures

Different measures can be used to assess the classifier performance. In this paper, the classifiers were evaluated using Accuracy, Recall, Precision, F-measures, AUC, and ROC Curve. These measures are calculated from the confusion matrix elements that consist of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP and TN determine if the classifier is working correctly, while FP and FN determine if the classifier is working in a bad way [18].

Accuracy measure identifies the number of correctly classified instances. The accuracy formula is represented in Eq. (1)

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

The precision measure is calculated by finding the percentage of instances classified as positive. The precision formula is represented in Eq. (2)

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

The recall is calculated by finding the percentage of positive instances classified as positive. The recall formula is represented in Eq. (3)

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

F-measure is a consistent mean of recall and precision. The F-measure formula is represented in Eq. (4) [19, 20]

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

Receiver Operating Characteristic (ROC) Curve is a useful tool to visualize the difference between classification models. The trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) is illustrated in ROC Curve. AUC determines the ability of the classifier to distinguish between the classes. A high value of AUC means a high ability to distinguish between positive and negative classes [21, 22]. Multiple studies prefer to use AUC measure in evaluating classifiers than accuracy measure [23, 24, 25]. We will use all these measures to evaluate the results in this study.

4.2 Analysis and Discussion of the Churn Prediction Model

The Deep learning model performance was compared to other classification algorithms as shown in Table 2. The comparison shows that the deep learning model obtained better results than Random Forrest, Decision Tree, and Gaussian NB in terms of precision, accuracy, recall, and F-measure.

Figure 2 demonstrates that the Deep Learning model has achieved better results and correctly classified the data with 91.98% accuracy. While Decision Tree and Gaussian NB have low accuracy rates of 67.82% and 70.18, respectively. On the other hand, Random Forrest performs better with an 80.71% accuracy rate. The performance of the different classifiers was also measured using AUC. The deep learning model surpassed other models with 0.917 AUC. In contrast, Random Forrest attained 0.770 AUC. Gaussian NB and Decision Tree achieved 0.655 AUC and 0.612 AUC respectively as illustrated in Figure 3.

Table 2- Performance measures for various Classifiers for churn Prediction

Prediction Model	Precision %	Accuracy %	Recall %	F-Measure %
Deep Learning	91.63	91.98	91.90	91.75
Random Forrest	83.51	80.71	77.66	78.65
Gaussian NB	72.72	70.18	65.60	65.25
Decision Tree	78.49	67.82	61.31	58.21

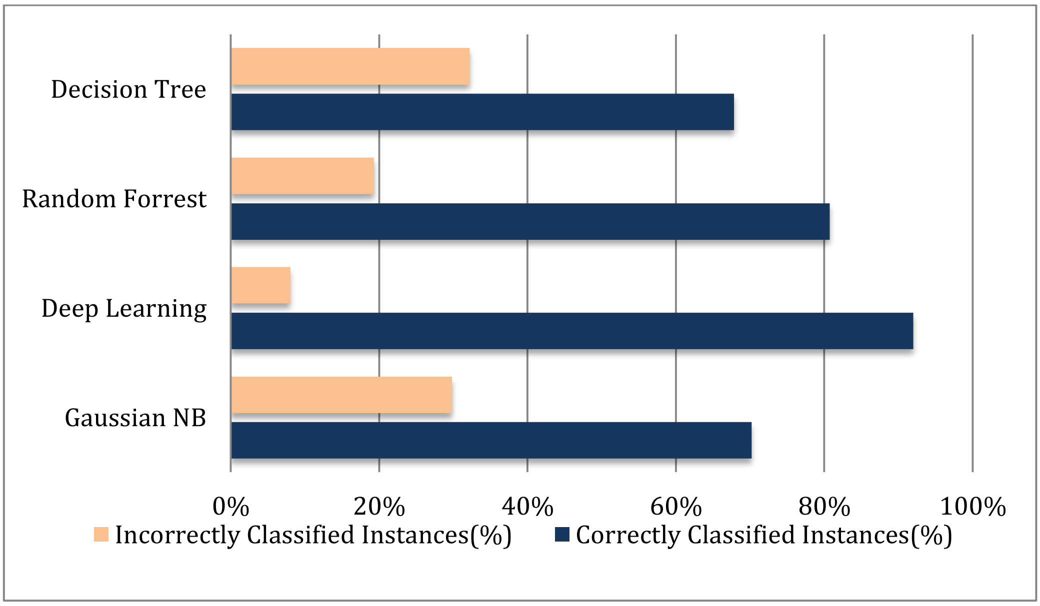

Figure 2-Accuracy Performance of Classification Algorithm

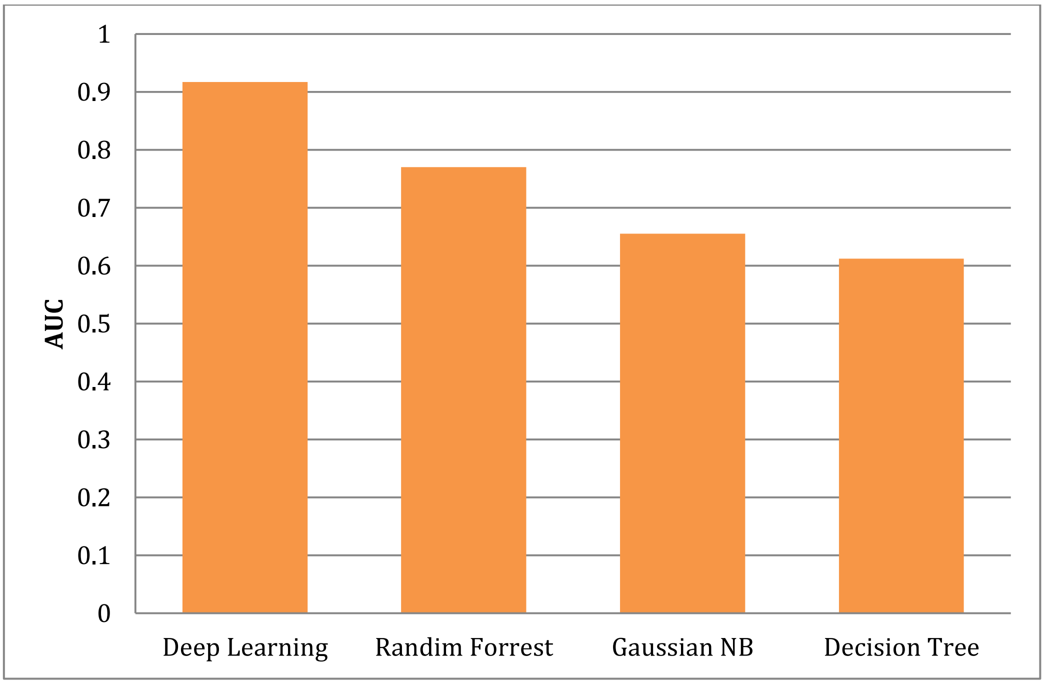

Figure 3-AUC Score of Various Classifiers

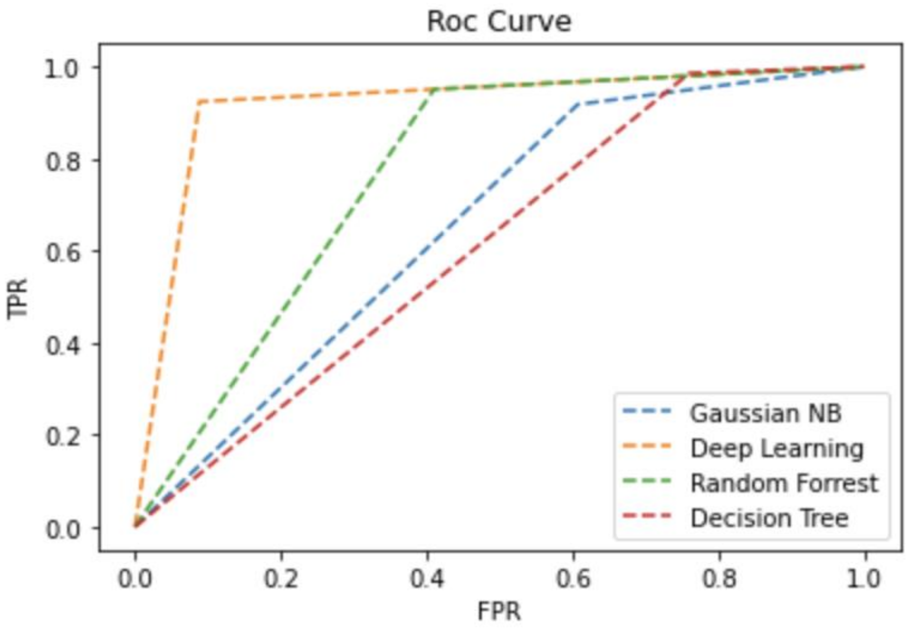

Figure 4-ROC Curve of Various Classifiers

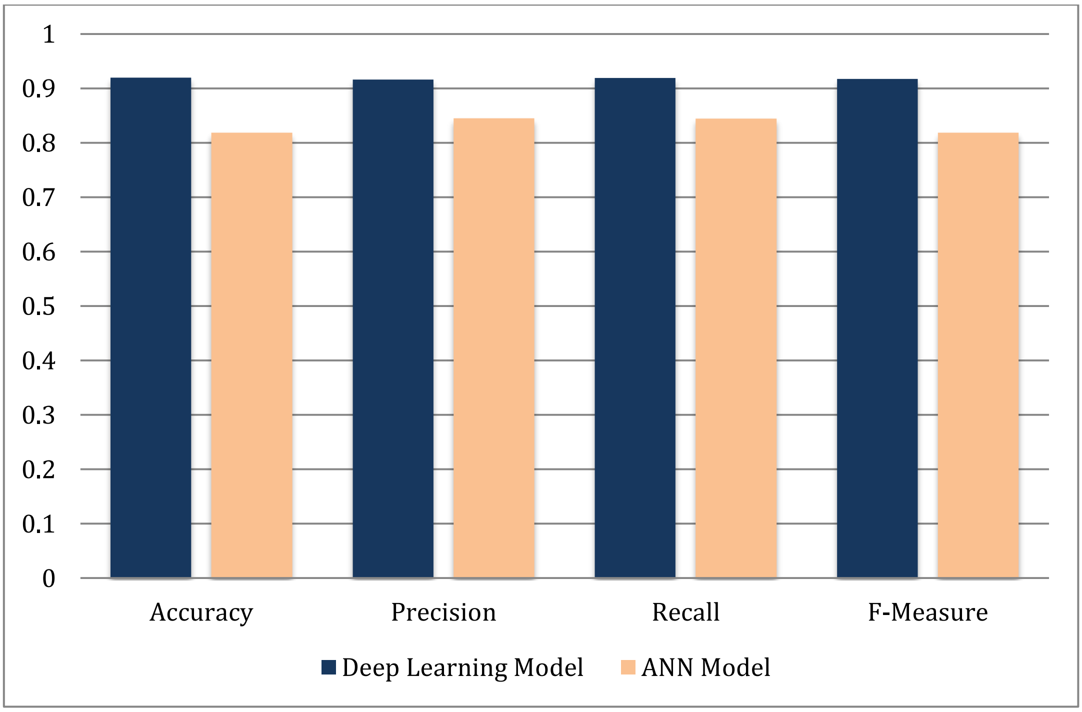

Figure 5-Performance of Deep Learning model against ANN model

For more demonstration of the models' performance, Figure 4 depicts the ROC curve of the various classifiers. From the figure, one can see that the Deep Learning model achieves high TPR in comparison to the other models.

The deep learning model performance was also compared to an ANN model that consists of 3 hidden layers that use Relu as an activation function and an output layer that uses Sigmoid as an activation function. The deep learning model achieves better results in terms of accuracy, precision, recall, and F-measure, as shown in Figure 5. The telecom operators always search for the models that achieve higher accuracy rate since the increase in accuracy means an increase in the company's revenue and a reduction of the cost.

5. Conclusion and Future Work

The competition between telecom operators has forced the companies to develop plans and offers to prevent customers from leaving their company. Developing an accurate model that can predict churn users based on their behavior is one of the most important aspects that decision-makers and project managers focus on due to the impact of churn subscribers on the company's growth. The main contribution of this paper is to develop a Deep Learning model that achieves a high accuracy rate of 91% compared to other traditional prediction models such as Decision Tree, Random Forrest, and Gaussian NB. The proposed model also outperforms in terms of precision, recall, F-measure, and AUC. The deep learning model works on high dimensional big data dataset without causing an overfitting issue in contrast to other models. Future work will focus on building a wisdom model of churn prediction that achieves a higher accuracy rate.

References

[1] M. Mozer, R. H. Wolniewicz, D. B. Grimes, E. Johnson, H. Kaushansky, and Others, "Churn reduction in the wireless industry," in *NIPS*, 1999, pp. 935–941.

[2] C.-S. Lin, G.-H. Tzeng, and Y.-C. Chin, "Combined rough set theory and flow network graph to predict customer churn in credit card accounts," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 8–15, Jan. 2011.

[3] I. Khan, I. Usman, T. Usman, G. U. Rehman, and A. U. Rehman, "Intelligent churn prediction for telecommunication industry," *International Journal of Innovation and Applied Studies*, vol. 4, no. 1, pp. 165–170, 2013.

[4] Lian Yan, R. H. Wolniewicz, and R. Dodier, "Predicting customer behavior in telecommunications," *IEEE Intell. Syst.*, vol. 19, no. 2, pp. 50–58, Mar. 2004.

[5] W. Verbeke, D. Martens, and B. Baesens, "Social network analysis for customer churn prediction," *Appl. Soft Comput.*, vol. 14, pp. 431–446, Jan. 2014.

[6] R. A. Soeini and K. V. Rodpysh, "Applying data mining to insurance customer churn management," *International Proceedings of Computer Science*, 2012, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1045.4033&rep=rep1&type=pdf.

[7] A. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 4, pp. 431–448, Oct. 2018.

[8] P. Paliwal and D. Kumar, "ABC Based Neural Network Approach for Churn Prediction in Telecommunication Sector," in *Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 2*, 2018, pp. 343–349.

[9] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019.

[10] Y. Khan, S. Shafiq, A. Naeem, and S. Hussain, "Customers churn prediction using artificial neural networks (ANN) in telecom industry," 2019, [Online]. Available: https://www.researchgate.net/profile/Mohammad_Alshammari/publication/336266496_Design_and_Learning_Effectiveness_Evaluation_of_Gamification_in_e-Learning_Systems/links/5daee953a6fdccc99d92b461/Design-and-Learning-Effectiveness-Evaluation-of-Gamification-in-e-Learning-Systems.pdf#page=142.

[11] J. Vijaya and E. Sivasankar, "Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector," *Computing*, vol. 100, no. 8, pp. 839–860, Aug. 2018.

[12] A. Idris, A. Iftikhar, and Z. ur Rehman, "Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling," *Cluster Comput.*, vol. 22, no. 3, pp. 7241–7255, May 2019.

[13] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting," *IEEE Trans. Industr. Inform.*, vol. 10, no. 2, pp. 1659–1665, May 2014.

[14] K. Dahiya and S. Bhatia, “Customer churn analysis in telecom industry,” in *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, Sep. 2015, pp. 1–6.

[15] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, “Customer churn prediction in telecommunication industry using data certainty,” *Journal of Business Research*, vol. 94. pp. 290–301, 2019, doi: 10.1016/j.jbusres.2018.03.003.

[16] A. A. Q. Ahmed and D. Maheswari, “An enhanced ensemble classifier for telecom churn prediction using cost based uplift modelling,” *Int. J. Inform. Technol.*, vol. 11, no. 2, pp. 381–391, Jun. 2019.

[17] A. Sharma and P. K. Panigrahi, “A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services,” *International Journal of Computer Applications*, vol. 27, no. 11. pp. 26–31, 2011, doi: 10.5120/3344-4605.

[18] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Elsevier, 2011.

[19] D. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, “Gene Selection and Classification of Microarray Data Using Convolutional Neural Network,” in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, Oct. 2018, pp. 145–150.

[20] S. Sharma and S. K. Srivastava, “Feature Based Performance Evaluation of Support Vector Machine on Binary Classification,” in *2016 Second International Conference on Computational Intelligence Communication Technology (CICT)*, Feb. 2016, pp. 170–173.

[21] D. Brzezinski and J. Stefanowski, “Prequential AUC: properties of the area under the ROC curve for data streams with concept drift,” *Knowl. Inf. Syst.*, vol. 52, no. 2, pp. 531–562, Aug. 2017.

[22] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning*. Springer Science & Business Media, 2011.

[23] F. Provost, T. Fawcett, and R. Kohavi, “The case against accuracy estimation for comparing induction algorithms 1998.”

[24] Jin Huang and C. X. Ling, “Using AUC and accuracy in evaluating learning algorithms,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, Mar. 2005.

[25] B.-G. Hu and W.-M. Dong, “A study on cost behaviors of binary classification measures in class-imbalanced problems,” *arXiv preprint arXiv:1403. 7100*, 2014.