International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 2, April 2024, pp. 2135~2144
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i2.pp2135-2144

Predicting churn with filter-based techniques and deep learning

Vivian Quek Jia Yi, Pang Ying Han, Lim Zheng You, Ooi Shih Yin, Wee How Khoh
Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia

Article Info

Article history:

Received Jul 10, 2023
Revised Oct 6, 2023
Accepted Nov 29, 2023

Keywords:

Artificial neural network
Attribute selection analysis
Churn prediction
Deep learning
Filter methods

ABSTRACT

Customer churn prediction is of utmost importance in the telecommunications industry. Retaining customers through effective churn prevention strategies proves to be more cost-efficient. In this study, attribute selection analysis and deep learning are integrated to develop a customer churn prediction model to improve performance while reducing feature dimensions. The study includes the analysis of customer data attributes, exploratory data analysis, and data preprocessing for data quality enhancement. Next, significant features are selected using two attribute selection techniques, which are chi-square and analysis of variance (ANOVA). The selected features are fed into an artificial neural network (ANN) model for analysis and prediction. To enhance prediction performance and stability, a learning rate scheduler is deployed. Implementing the learning rate scheduler in the model can help prevent overfitting and enhance convergence speed. By dynamically adjusting the learning rate during the training process, the scheduler ensures that the model optimally adapts to the data while avoiding overfitting. The proposed model is evaluated using the Cell2Cell telecom database, and the results demonstrate that the proposed model exhibits a promising performance, showcasing its potential as an effective churn prediction solution in the telecommunications industry.

Corresponding Author:

Pang Ying Han
Faculty of Information Science and Technology, Multimedia University
Ayer Keroh Lama Street, 75450 Ayer Keroh, Melaka, Malaysia
Email: yhpang@mmu.edu.my

1. INTRODUCTION

Customers are the backbone of any business, and their loyalty is critical to its success. In the rapidly growing and competitive telecommunications industry, customer churn is a major concern because retaining existing customers is more cost-effective than acquiring new customers. The costs of acquiring new customers associated with marketing, discounts, and training new staff are substantial [1]. Customer churn is an important indicator for any company that offers subscription-based services, and it's of utmost importance in the telecommunications industry [2]. It can be caused by a variety of factors, including poor customer service or rival competitors providing more appealing discounts. In order to reduce the rate of churn among existing customers, companies must realize the importance of developing long-term connections with their customers rather than simply focusing on gaining new ones. A high churn rate indicates that clients are unsatisfied with the service and that the supplier is failing to satisfy their expectations [3]. Hence, telecommunication companies should prioritize engaging in initiatives that reduce churn and retain customers. For example, telecommunication companies might implement innovative strategies to retain clients during the pandemic, such as integrating online business on their e-commerce platform or requiring virtual appointments to regulate customer flow [4]. The epidemic has sped the move to digital channels, and to compete, telecommunications companies must adapt to this new reality [5].

Happy customers are more likely to stay and recommend the company to others. Therefore, this will result in higher revenue and growth. Companies should strive to improve their customer service, provide more value-added services, and build trust and loyalty with their customers to prevent them from churning. However, if these strategies are applied to all the existing customers, the costs are high and financially inefficient too.

This study presents an approach to address the challenge of customer churn prediction. The incorporation of a filter-based attribute selection technique and deep learning in the proposed model enables effective churn detection and prevention, yielding a promising F1 score. The high F1 score implies that the model is able to accurately detect customers who are likely to churn while minimizing both false negatives and positives. This signifies that the proposed model can be a valuable tool for businesses, assisting them in determining potential churn customers. With this, the companies can offer extraordinary services to those potential churn customers to prevent them from churning. Besides that, this approach also serves to alleviate the costly mistake of misclassifying loyal customers as potential churners. In this study, two attribute selection techniques, namely the analysis of variance (ANOVA) test and chi-square, are employed to identify the most significant features in the churn prediction model. These techniques enable the model to focus on the attributes that have the strongest association with churn, improving the model's accuracy [6]. In the proposed model, data preprocessing and exploratory data analysis (EDA) are implemented to better comprehend consumer data. To enhance the quality of customer data and optimize model performance, a series of pre-processing processes are applied. These involve various steps aimed at cleaning, transforming, and preparing the data for churn prediction. Two attribute selection approaches, namely ANOVA and chi-square, are employed to investigate and select key features from the pre-processed data. These methods aid in determining the most informative and adequate features for churn prediction. Finally, the chosen features are input into an artificial neural network model (ANN) for churn prediction. The ANN model leverages the power of deep learning algorithms to capture complex patterns and relationships within customer data. In this study, the Cell2Cell dataset is used for predicting customer churn. The Cell2Cell dataset contains diverse customer data, including customer service interactions, usage patterns, demographic information, and other variables. This dataset enables a comprehensive analysis of customer behavior, offering valuable insights into the factors that contribute to churn in the telecommunications industry [7]. This dataset is available at this URL link: https://www.kaggle.com/datasets/jpacse/datasets-for-churn-telecom [8]. The telecommunications industry experiences a high churn rate. Companies should prioritize customer retention so that they can build a long-term customer base and increase income. As a result, by prioritizing customer retention, companies can establish a sustainable customer base and boost revenue over the long term [9].

A significant portion of papers utilizing the Cell2Cell dataset for customer churn prediction predominantly favored machine learning algorithms [10]. The utilization of deep learning models for churn prediction on the Cell2Cell dataset remains relatively scarce. Hence, this study focuses on studying how deep learning can help enhance prediction performance rather than using traditional machine learning algorithms. Deep learning is widely regarded as a powerful technique for modelling complex data patterns due to its ability to identify complex correlations and extract high-level features from raw data. The deep learning models can be categorized according to the primary purpose of the study, such as classification, feature extraction, prediction, or expression [11]. This approach excels at predicting customer behavior since it can automatically learn intricate and nonlinear correlations between the numerous aspects that contribute to customer churn. Moreover, deep learning techniques can detect possible churn by recognizing trends in customer behavior, such as changes in purchase frequency or engagement. Additionally, because these models can be trained on large datasets, they can capture various representations of customer behavior and enhance the model's performance. Since the Cell2Cell dataset consists of a total of 71,047 instances and 58 attributes, it is considered a large dataset for predicting customer churn [12]. Research that was previously conducted stated that the Cell2Cell dataset is not an easy dataset for performing predictions due to its imbalanced churn rate [13]. Therefore, deep learning models are proposed for training with the Cell2Cell dataset due to their capability to discern intricate patterns and relationships within the data. Furthermore, deep learning models exhibit the ability to adapt to changes in customer behavior and leverage large volumes of data, making them effective for churn prediction. Hence, it is widely used to solve customer churn problems [14].

In light of this, this proposed system is based on an ANN model. The main contributions to this paper are: i) To address the imbalanced class issue in the Cell2Cell dataset, an oversampling technique, namely the synthetic minority oversampling technique (SMOTE), is applied. As a side note, SMOTE is only applied to the training set. This is to ensure the integrity of the model prediction process and facilitate a more accurate evaluation of the model's performance on unseen samples (i.e., the test set) [15]; ii) To identify attributes with a higher correlation while eliminating those that have a lower correlation. Since there are over 50 attributes in the Cell2Cell dataset, attribute selection analysis is performed. Attribute selection can help in selecting the most relevant and representative characteristics that can effectively predict customer churn behavior. By narrowing down the attribute set, attribute selection enhances the accuracy of the churn prediction model; and iii) To mitigate the risk of overfitting and improve training efficiency, a learning rate scheduler is incorporated into the proposed model. By dynamically adjusting the learning rate, the model becomes less prone to overfitting, ensuring better generalization performance on unseen data.

2. THE PROPOSED METHOD

In this work, the Cell2Cell dataset is used to train the proposed artificial neural network (ANN) model. As depicted in Figure 1, the proposed system architecture for churn prediction is implemented in five stages: data retrieval and exploratory data analysis, data preprocessing, attribute selection analysis, classification, and model evaluation. The details of the processes will be explained in the subsequent sections.

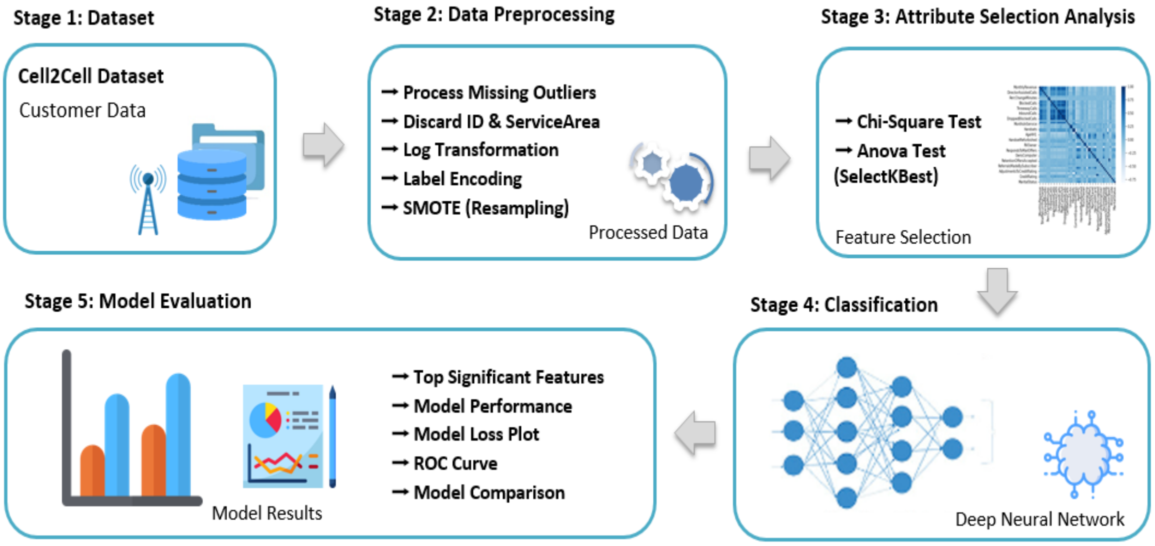


Figure 1. The proposed system architecture for churn prediction

2.1. Exploratory data analysis and data preprocessing

In order to ensure proper training of the proposed model, the data undergoes two essential stages: EDA and data preprocessing. EDA allows visualizing and comprehending the data through statistical techniques and graphical representations to provide insights for the subsequent preprocessing step. On the other hand, data is cleaned and transformed for enhanced quality and improved compatibility with the model. The Cell2Cell dataset is stored in a comma-separated values (CSV) file and displayed using the Pandas DataFrame. This data frame is a two-dimensional (2D) structure used to analyze labelled tabular data. EDA is performed to comprehend the data in the Cell2Cell dataset. The findings resulting from performing the EDA show that several strategies may be used to examine data for further processing and improve the performance of prediction models [16].

Besides that, data preprocessing is mandatory to clean up and adjust the problematic variables and improve the prediction results. First, there are several variables in the dataset consisting of outliers and missing values, i.e. roaming calls and monthly revenue. The np.nan method is applied to fill up the data by replacing missing variables with zero instead of eliminating null values, whereas the is null() method is used to determine missing values. By performing EDA, two attributes have been excluded from this study: i) the "CustomerID" attribute, which serves as a unique identifier for each client, and this attribute does not provide any meaningful insights into predicting churn behavior; and ii) the "ServiceArea" attribute, due to its potential to introduce excessive noise in the dataset [17]. This decision is primarily based on the observation that when the categorical variables are label encoded, the “ServiceArea” attribute would generate around 740 distinct category labels. The high number of labels can adversely impact the model's performance and introduce unnecessary complexity. Thus, to streamline the dataset and maintain model efficiency, the "ServiceArea" attribute is removed from further analysis.

Next, log transformation is applied to the numerical variables to reduce their skewness. The reason to reduce the skewness of the numerical variables is to normalize the distribution of the variables so that these variables can be normalized. In addition, to standardize the data format for the following procedures, some variables require data transformation. For instance, the categorical variables should undergo label encoding from string format to integer format before training the model [18]. Figure 2 demonstrates an example of label encoding for a categorical variable. The categorical variable “PrizmCode” with categories “Other”, “Suburban”, "Town,” and “Rural” is encoded as “Other”=0, “Suburban”=1, “Town”=2, and “Rural” = 3.

Class data imbalance is a common concern in binary classification tasks. It refers to a scenario where the number of instances in one class significantly outweighs the other, resulting in an imbalanced distribution. In customer churn prediction, the churn class (customers who churn) is significantly smaller than the non-churn class (customers who do not churn). If the data is imbalanced, it may not accurately reflect the issue in the real world. Furthermore, most of the literature is applied to a small number of datasets. Working with only a small number of datasets may hinder the exploration of diverse data patterns and characteristics, potentially limiting the generalizability and effectiveness of the models [19], [20]. The Cell2Cell dataset is acknowledged to exhibit a significant class imbalance issue, with ~70% of non-churn samples (i.e., the majority class) and ~20% of churn samples (i.e., the minority class).

To balance the dataset's churn rate, SMOTE is applied for data sampling. SMOTE is an oversampling strategy. Data sampling balances the dataset and gives a more descriptive sample to the model by boosting the minority class’s representation. A balanced dataset is necessary since it prevents the model from being biased towards one class, allowing accurate predictions [21]. By implementing SMOTE, it generates synthetic minority class observations by interpolating between existing minority class observations. This is achieved by selecting an observation at random from a minority class and then determining its k nearest neighbors in the feature space. Then, one of the neighbors is chosen at random, and a new observation is created by interpolating between the chosen observation and the randomly chosen neighbor. After applying SMOTE, the class distribution was rebalanced, resulting in an equal number of churn and non-churn samples.

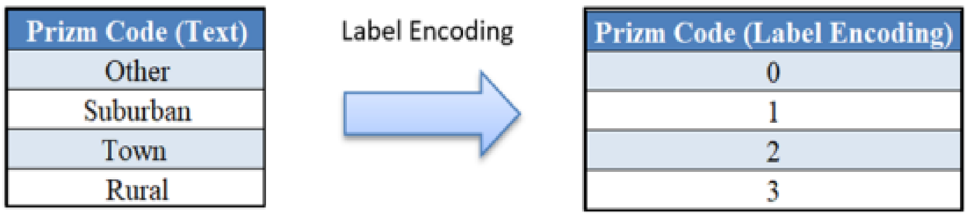


Figure 2. Label encoding

2.2. Attribute selection analysis

When a dataset contains numerous attributes, the model's performance may be hindered due to the abundance of information. The presence of numerous attributes in a dataset can introduce a high level of noise when fitting it into a model, thereby affecting its performance. Studies have shown that attribute selection techniques help improve model performance by reducing feature dimensions. Thus, performing attribute selection analysis before training the model for customer churn prediction is crucial to mitigate the impact of noise and ensure optimal performance. However, discovering the most effective features for customer churn prediction can be a challenging task due to the complexity and variability of the dataset [22]. Employing attribute selection analysis can serve as a viable solution to identify the most relevant features for customer churn prediction while eliminating the less correlated attributes. In essence, attribute selection aids in determining the most indicative characteristics that result in better results for churn prediction.

Two filter-based attribute selection techniques are studied: the ANOVA test method and the chi-square test. The ANOVA test is a feature selection method that evaluates feature scores statistically [23]. The ANOVA statistical test selects the top d features with the highest scores and excludes those unselected features. This strategy is to reduce the number of dimensions in the data while retaining the most significant features, making it a popular choice in deep learning to improve classification performance. The Chi-Square test is a statistical method used to evaluate the degree of independence between two variables in a dataset. Based on actual and anticipated frequencies, it assesses if there is a substantial link or relationship between the variables. It identifies highly related sample data before removing redundancy with a minimum redundancy algorithm [24]. A lower chi-square value implies that the variables are less likely to be related, whereas a higher score shows that the variables are strongly associated. In this study, the top 40 significant features are selected.

2.3. Classification model

As aforementioned, the proposed model for the task of churn prediction is the artificial neural network (ANN) model. An ANN model is one of the most commonly used deep learning models for addressing complex prediction problems by developing a neural network that imitates the neuron system in the human brain. The ANN model can be formulated as (1):

$$z = f(b + x \cdot w) = f(\textstyle\sum_{i=1}^{n} x_i w_i + b) \tag{1}$$

where
w : denotes the vector of weights,
x : the vector of inputs,
b : the bias.

In this study, the proposed model is a multi-layer ANN model that incorporates regularization techniques, such as L2 regularization, batch normalization, and dropout, to minimize the risk of overfitting and improve generalization performance. The first layer is a dense layer with 128 neurons and a rectified linear unit (ReLU) activation function to introduce nonlinearity to the output of the layer. It also comprises L2 regularization to prevent overfitting. Next, a batch normalization layer is added to help normalize the outputs of the previous layer for model stability and training efficiency. To further promote model generalization and mitigate overfitting, a dropout layer is included to randomly deactivate a fraction of neurons during training. The same patterns of dense, batch normalization, and dropout layers are repeated twice, reducing the number of neurons in each subsequent layer (i.e., 64 and 32). This is to create a deeper network architecture that can capture more abstract features, extracting meaningful data patterns. Lastly, the final layer is a dense layer to produce the output prediction for binary classification, i.e., churn or non-churn. To optimize prediction performance and improve model stability, a dynamic learning rate scheduler is implemented. This learning rate scheduler helps prevent model overfitting and accelerates convergence with dynamic learning halved every 10 epochs. Figure 3 illustrates the overview of the proposed ANN model.

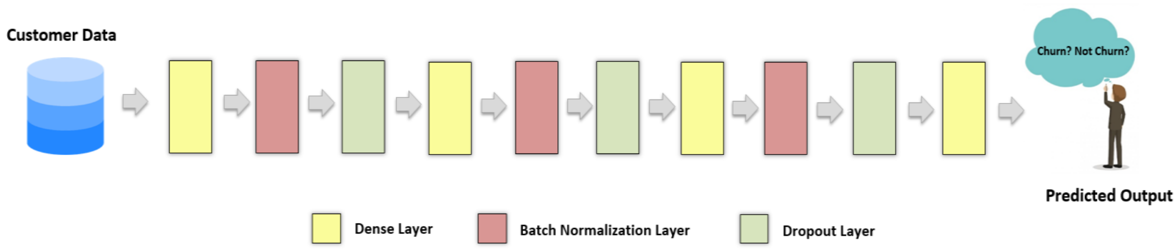


Figure 3. The overview of the proposed ANN model architecture

3. RESULTS AND DISCUSSION

3.1. Experimental setup and performance metrics

In this study, the experiments are conducted using the Jupyter Notebook and Keras libraries. The proposed ANN-based customer churn prediction model is assessed using a training-validation-test split-testing strategy. The Cell2Cell dataset is separated into three portions: i) a training set for model training; ii) a validation set to fine-tune the model hyperparameters and make decisions regarding the model's optimization; and iii) a testing set for obtaining the final model evaluation performance after the model training is completed.

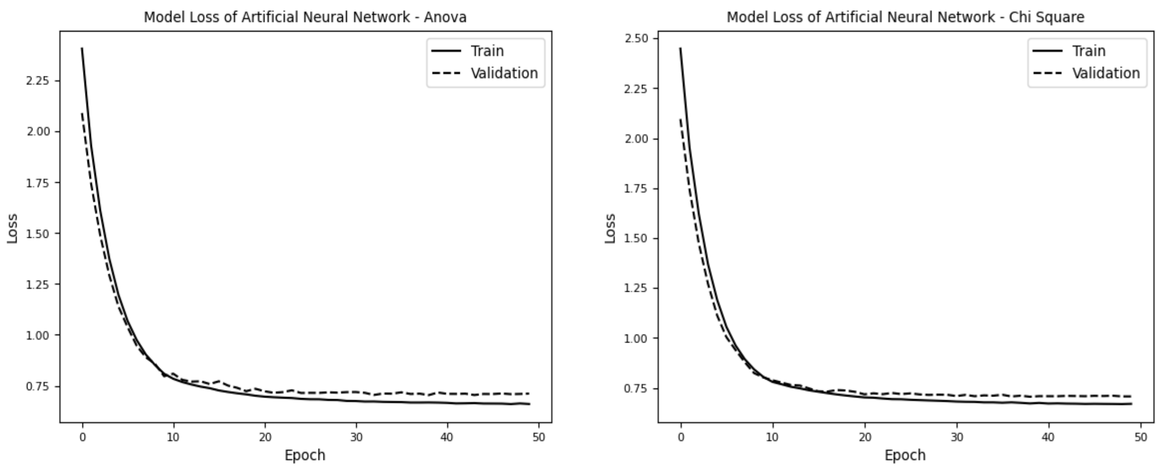
3.2. Model performance analysis

Attribute selection analysis is conducted before model training to identify the most significant attributes as inputs for the model. Table 1 shows the performance (in terms of accuracy, precision, recall, and F1 score) of the proposed churn prediction model with different attribute selection analyses. From the empirical results, we can observe that the proposed system using the ANOVA test as attribute selection analysis (called ANN-Ano) performs better than that using the chi-square test (known as ANN-Chi). The former model achieves an accuracy score of 58.37%, a precision score of 66.29%, a recall score of 58.37%, and an F1 score of 60.32%, whereas the latter model obtains an accuracy score of 55.59%, a precision score of 58.04%, a recall score of 55.59%, and an F1 score of 53.54%. Moreover, the experimental results show that the proposed model using a full feature set for classification (called ANN-Full) attains an accuracy score of 57.84%, a precision score of 66.15%, a recall score of 57.84%, and an F1 score of 59.82%, respectively. From the empirical results, we can deduce that the adoption of an adequate attribute selection technique for extracting those significant features in a model is helpful. Despite using a smaller set of features (i.e., 40 features in this case), ANN-Ano achieves a slightly higher accuracy and F1 score compared with ANN-Full. This suggests that the selected subset of features in ANN-Ano contains the most relevant attributes for churn prediction. Furthermore, using a subset of features can reduce dimensionality, mitigate the risk of overfitting, and enhance computational efficiency. These advantages make ANN-Ano a more efficient choice.

Table 1. Classification performance of the proposed ANN model

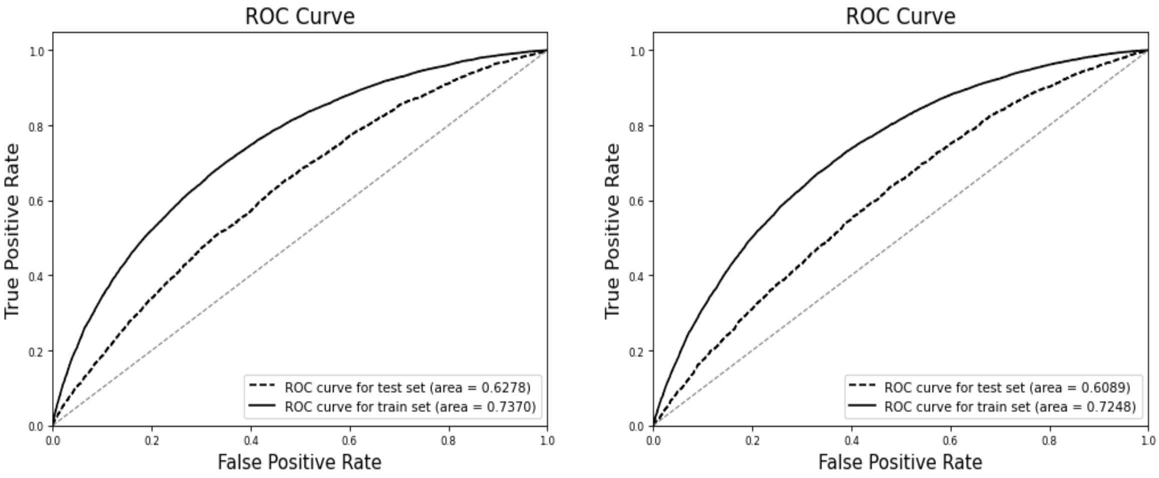
Features	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Using a full feature set (called ANN-Full)	57.84	66.15	57.84	59.82
Using the top 40 features from ANOVA test (called ANN-Ano)	58.37	66.29	58.37	60.32
Using the top 40 features from chi-square test (called ANN-Chi)	55.59	58.04	55.59	53.54

Figures 4 and 5 show the loss function train plot and receiver operating characteristic (ROC) curve, respectively, for better illustration. To be more specific, Figure 4(a) and Figure 4(b) illustrate the loss function plots of the ANN-Ano model and the ANN-Chi model, respectively. On the other hand, the ROC curves of the ANN-Ano model and the ANN-Chi model are shown in Figure 5(a) and Figure 5(b), respectively. By comparing the training and validation losses, the model's generalization ability can be assessed. We could observe that both losses decrease and converge to low values. The decreasing training loss and the decreasing validation loss indicate that the proposed model is learning effectively. Additionally, we also notice that the area under the ROC curve for ANN-Ano is higher than that of ANN-Chi. This indicates that the former model can effectively capture the underlying patterns and dependencies in the data more accurately and make more accurate predictions. Consequently, the model demonstrates stronger predictive accuracy and a greater capacity to distinguish between the two classes, namely churn and non-churn in this case.



(a) (b)

Figure 4. Loss function plot of the ANN model (a) ANOVA (ANN-Ano) and (b) chi-square (ANN-Chi)



(a) (b)

Figure 5. ROC curve of the proposed ANN model (a) ANOVA (ANN-Ano) and (b) chi-square (ANN-Chi)

3.3. Comparison and discussions

The performance of the proposed ANN model is compared with the performance of other models that are using the same database. Table 2 shows the performance results of the proposed model and the other existing models. When evaluating the performance of a deep neural network model for predicting customer churn rate, precision, recall rate, and F1 score are often considered more important than accuracy. Accuracy alone can be misleading in such cases because a model can achieve high accuracy by simply predicting the majority class (non-churn) for all instances. This would result in high accuracy but fail to capture the churn cases that are crucial for businesses. The primary goal in churn prediction is to correctly identify as many churn cases as possible. Recall rate, also known as sensitivity or true positive rate, measures the proportion of actual churn cases that are correctly predicted as churn. It reflects the model's ability to capture the positive class accurately. A high recall rate implies that the model is effective at identifying customers who are likely to churn, which is critical for businesses to take proactive measures to retain those customers. Precision measures the proportion of predicted churn cases that are true churn cases. It reflects the model's ability to avoid false positives, i.e., not classifying non-churn customers as churn. The F1 score combines both precision and recall into a single metric, providing a balanced evaluation of the model's performance. Hence, precision, recall rate, and F1 score provide a better understanding of how well the model identifies the positive class (churn).

Table 2. Model comparison of the existing approaches in the Cell2Cell dataset

Approach	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Proposed ANN model (ANN-Ano)	58.37	66.29	58.37	60.32
ANN model without SOM* [25]	57.59	NA	59.45	58.37
Baseline classifier on the default setting (overall)* [26]	57	56	55	54
Transfer learning (TL-DeepE)* [27]	68.16	NA	NA	NA
Naïve Bayes* [15]	69.17	37.56	10.83	16.81
Undersampling + Decision tree* [28]	NA	2.16	51.97	NA
Random forest* [29]	63.09	34.67	27.05	29.34

** Results are extracted from the respective authors and papers*

The empirical results indicate that the proposed ANN model, i.e., ANN-Ano, outperforms the other approaches in terms of precision and F1 score. This improvement can be attributed to the effective utilization of attribute selection, specifically employing the ANOVA SelectKBest method to identify the top 40 significant features. These carefully chosen features have significantly contributed to enhancing the performance of the ANN model, resulting in superior precision and F1 scores. The model's improved precision and F1 score suggest that it excels in correctly identifying positive instances while minimizing false positives and false negatives. This ability is valuable in customer churn prediction, where accurate predictions and minimizing errors are crucial. On the other hand, we could notice that the ANN without a self-organizing map (SOM) achieves a slightly better recall rate than the proposed ANN model. However, the performance in terms of the F1 score is slightly poorer than that of the proposed ANN model, achieving a 58.37% F1 score. This could be due to the insufficient training sample to generalize and optimize the model learning. Besides, naive Bayes, decision tree, and random forest may not perform well in forecasting customer churn in the Cell2Cell dataset. This may be due to the independence assumption of naive Bayes and the limitations of random forest in capturing complex patterns and dependencies. In addition, the utilization of a learning rate scheduler during training in the proposed ANN model results in a remarkably short training time, with less than 10 seconds per epoch. By incorporating the learning rate scheduler, the model dynamically adjusts the learning rate during training, ensuring optimal progress in updating the model's parameters. This adaptive learning rate strategy enhances the convergence speed and the model's ability to find an optimal solution within a shorter time frame.

4. CONCLUSION

This paper proposes an ANN deep learning model for predicting customer behavior. The proposed ANN model is tested with the Cell2Cell dataset, which is a real-world telecommunications database. To enhance the data quality, customer data is subjected to exploratory data analysis and data preprocessing. Next, relevant characteristics are selected using two attribute selection techniques: ANOVA and chi-square tests. The proposed model is fed with the most significant features for data analysis and classification. From the empirical results, the model with the ANOVA test (i.e., ANN-Ano) achieves the highest F1 score of 60.32%, followed by the model with the chi-square test, i.e., ANN-Chi (53.54%), and the model with the entire feature set, i.e., ANN-Full (59.82%). Hence, the ANOVA test is determined to be a suitable attribute selection technique compared to the chi-square test for selecting the best representative features from the Cell2Cell dataset to predict churn rate. Besides that, the proposed ANN model achieves the best performance in terms of precision and F1 score and the second-best recall rate compared to the other ANN and machine learning models. In addition, the proposed ANN model has a training time of fewer than 10 seconds per epoch by using the learning rate scheduler. In summary, the incorporation of appropriate exploratory data analysis, data preprocessing, ANOVA-based attribute selection and a customized ANN model demonstrates the efficacy of the proposed system with the highest F1 score among the other compared approaches in churn prediction task. This good performance underlines an efficient methodology for companies to enhance customer retention. This high F1 score implies that the proposed system is highly effective in detecting potential churn customers while diminishing false positives and negatives. With that, it helps the companies in resource allocation and targeted retention strategies. Furthermore, this approach can serve as a valuable basis for predictive analytics across different sectors, including banking, insurance companies and e-commerce, for customer behavior analysis.

ACKNOWLEDGEMENTS

This research is supported by MMU Postdoctoral Research Fellow Grant, MMUI/230003.

REFERENCES

[1] K. Matuszelański and K. Kopczewska, “Customer churn in retail e-commerce business: Spatial and machine learning approach,” *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 1, pp. 165–198, 2022.
[2] W. H. Khoh, Y. H. Pang, S. Y. Ooi, L.-Y.-K. Wang, and Q. W. Poh, “Predictive churn modeling for sustainable business in the telecommunication industry: Optimized weighted ensemble machine learning,” *Sustainability*, vol. 15, no. 11, Art. no. 8631, 2023.
[3] C. P. Johny and P. P. Mathai, “Customer churn prediction: a survey,” *International Journal of Advanced Research in Computer Science*, vol. 8, no. 5, pp. 2178–2181, 2017.
[4] J. Roe, G. Brady, L. Emery, S. Walker, I. de Brito Ashurst, and D. Mowbray-Pape, “Co-design of new post-covid oncology rehabilitation services offers a model for the future,” *BMJ*, vol. 379. British Medical Journal Publishing Group, 2022.
[5] N. Mustafa, L. S. Ling, and S. F. A. Razak, “Customer churn prediction for telecommunication industry: A Malaysian Case Study,” *F1000Research*, vol. 10, 2021.
[6] M. Sadeghi, M. N. Dehkordi, B. Barekatain, and N. Khani, “Improve customer churn prediction through the proposed PCA-PSO-K means algorithm in the communication industry,” *The Journal of Supercomputing*, vol. 79, no. 6, pp. 6871–6888, 2023.
[7] I. U. Khan and N. L. S. Cherukupalli, “Research on the analytics of traffic pumping in telecommunications via data science using rehabilitated frog leaping algorithm,” *International Journal of Data Informatics and Intelligent Computing*, vol. 2, no. 1, pp. 51–61, 2023.
[8] Pamina, “Telecom churn (cell2cell),” *Kaggle*. https://www.kaggle.com/datasets/jpacse/datasets-for-churn-telecom (accessed Jan. 10, 2023).
[9] M. Zhao, Q. Zeng, M. Chang, Q. Tong, and J. Su, “A prediction model of customer churn considering customer value: an empirical research of telecom industry in China,” *Discrete Dynamics in Nature and Society*, vol. 2021, pp. 1–12, 2021.
[10] S. W. Fujo, S. Subramanian, and M. A. Khder, “Customer churn prediction in telecommunication industry using deep learning,” *Information Sciences Letters*, vol. 11, no. 1, p. 24, 2022.

[11] K. Albulayhi, Q. Abu Al-Haija, S. A. Alsuhibany, A. A. Jillepalli, M. Ashrafuzzaman, and F. T. Sheldon, "IoT intrusion detection using machine learning with a novel high performing feature selection method," *Applied Sciences*, vol. 12, no. 10, Art. no. 5015, 2022.

[12] P. Singh, S. Bhardwaj, S. Dixit, R. N. Shaw, and A. Ghosh, "Development of prediction models to determine compressive strength and workability of sustainable concrete with ANN," in *Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2021*, 2021, pp. 753–769.

[13] N. N. Nguyen and A. T. Duong, "Comparison of two main approaches for handling imbalanced data in churn prediction problem," *Journal of advances in information technology*, vol. 12, no. 1, 2021.

[14] A. Dalli, "Impact of hyperparameters on deep learning model for customer churn prediction in telecommunication sector," *Mathematical Problems in Engineering*, vol. 2022, 2022.

[15] M. Waqar, H. Dawood, H. Dawood, N. Majeed, A. Banjar, and R. Alharbey, "An efficient SMOTE-based deep learning model for heart attack prediction," *Scientific Programming*, vol. 2021, pp. 1–12, 2021.

[16] K. Zheng, "Identifying churning employees: Machine learning algorithms from an unbalanced data perspective," in *ACM International Conference Proceeding Series*, 2022, pp. 14–22, doi: 10.1145/3568199.3568202.

[17] H. Jain, A. Khunteta, and S. Srivastava, "Telecom churn prediction using an ensemble approach with feature engineering and importance," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 3, pp. 22–33, 2022.

[18] S. Khotijah, "Churn prediction," *Kaggle*. https://www.kaggle.com/khotijahs1/churn-prediction (accessed Jun. 25, 2022).

[19] M. H. Seid and M. M. Woldeyohannis, "Customer churn prediction using machine learning: Commercial bank of Ethiopia," in *2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, 2022, pp. 1–6.

[20] A. Muneer, R. F. Ali, A. Alghamdi, S. M. Taib, A. Almaghthawi, and E. A. A. Ghaleb, "Predicting customers churning in banking industry: A machine learning approach," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 26, no. 1, pp. 539–549, 2022.

[21] S. Pokhrel, R. Abbas, and B. Aryal, "IoT security: botnet detection in IoT using machine learning," *arXiv preprint arXiv:2104.02231*, 2021.

[22] Y. T. Naing, M. Raheem, and N. K. Batcha, "Feature selection for customer churn prediction: A review on the methods & techniques applied in the telecom industry," in *IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics, ICDCECE 2022*, 2022, pp. 1–5, doi: 10.1109/ICDCECE53908.2022.9793315.

[23] K. Lazaros, S. Tasoulis, A. Vrahatis, and V. Plagianakos, "Feature selection for high dimensional data using supervised machine learning techniques," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 3891–3894.

[24] Y. Wang and C. Zhou, "Feature selection method based on chi-square test and minimum redundancy," in *Emerging Trends in Intelligent and Interactive Systems and Applications: Proceedings of the 5th International Conference on Intelligent, Interactive Systems and Applications (IISA2020)*, 2021, pp. 171–178.

[25] M. Fathian, Y. Hoseinpoor, and B. Minaei-Bidgoli, "Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods," *Kybernetes*, 2016.

[26] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naive Bayes," *Applied Soft Computing*, vol. 137, Art. no. 110103, 2023.

[27] U. Ahmed, A. Khan, S. H. Khan, A. Basit, I. U. Haq, and Y. S. Lee, "Transfer learning and meta classification based deep churn prediction system for telecom industry," *arXiv preprint arXiv:1901.06091*, 2019.

[28] W. Ma, G. Xia, S. Chen, and G. Li, "Customer churn prediction: balanced random forest with feature under-sampling," in *CAIBDA 2022; 2nd International Conference on Artificial Intelligence, Big Data and Algorithms*, 2022, pp. 1–7.

[29] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated churn prediction and customer segmentation framework for telco business," *IEEE Access*, vol. 9, pp. 62118–62136, 2021.

BIOGRAPHIES OF AUTHORS

Vivian Quek Jia Yi is currently pursuing her degree in artificial intelligence from Multimedia University, Malaysia. Her research interests include machine learning, deep learning, image processing and pattern recognition. She can be contacted at email: 1181202878@student.mmu.edu.my.

Pang Ying Han received her B.E. (Hons) degree in electronic engineering in the year 2002, Master of Science degree in the year 2005 and PhD degree in the year 2013 from Multimedia University. She is currently an associate professor in the Faculty of Information Science and Technology, at Multimedia University, Malaysia. Her research interests include human activity recognition, machine learning, deep learning and pattern recognition. She can be contacted at email: yhpang@mmu.edu.my.

Lim Zheng You received the Bachelor Degree of Engineering (Hons.) Electronics majoring in robotics and automation, Master of Engineering Science (M.Eng.Sc.), and PhD in engineering in 2016, 2019, and 2023 respectively from Multimedia University, Malaysia. He is currently a post-doctoral research fellow in Multimedia University. His research interests are robotics, automations, image processing, artificial intelligence, and electroencephalogram. He can be contacted at email: lim.zhengyou@mmu.edu.my.

Ooi Shih Yin received her Bachelor of Information Technology (Hons), Master of Science and PhD from the Multimedia University, Malaysia. She was a research fellow with Yonsei University, South Korea, 2018-2019. She is currently an associate professor in the Faculty of Information Science and Technology, at Multimedia University. Malaysia. Her research interests include image processing, computer vision, and machine learning. She can be contacted at email: syooi@mmu.edu.my.

Wee How Khoh received his Master of Science in information technology and Ph.D. degree from Multimedia University. He is also working as a senior lecturer in the Faculty of Information Science and Technology (FIST), at Multimedia University. His research interests are machine learning, biometrics authentication and security, and signal and image processing. He can be contacted at email: whkhoh@mmu.edu.my.