Published in AJSE, Vol:23, Issue: 1
Received on 22nd December 2022
Revised on 18th January 2024
Published on 25th April 2024

AIUB JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 1608 – 3679 (print) 2520 – 4890 (Online)
DOI: 10.53799/ajse.v23i1.612

# A Comparison of Customer Churn Vector Embedding Models with Deep Learning

Dinne Ratj, Tjeng Wawan Cenggoro, Namira Mufida Adien, Ni Putu Putri Ardhia Paramita, Nabila Putri Sina, Gregorius Natanael Elwirehardja and Bens Pardamean

***Abstract*—In the telecommunication industry, deep learning has been utilized for churn prediction. Some companies have used sophisticated deep learning techniques to predict churn, which yielded satisfactory results. However, future studies are still required to evaluate several deep learning mechanisms as only SoftMax Loss has been used so far. By comparing customer churn vector embedding models with several methods, including SoftMax Loss, Large Margin Cosine Loss, Semi-Supervised Learning, and a combination of Large Margin Cosine Loss and Semi-Supervised Learning. The use of Large Margin Cosine Loss has been proven in face recognition which can increase the discrimination between vectors embedding in different classes. Understanding how mixing unlabeled and labeled input might alter developing algorithms and learning behavior that benefit from this combination are the goals of semi-supervised learning. This approach encouraged feature discrimination in customer behavior and improved the model's overall accuracy. Large Margin Cosine Loss in this study achieved 83.74% of the F1 Score compared to other methods. It was further demonstrated that the produced vectors for churn prediction are discriminative by examining the cluster's similarity and the t-SNE plot. The t-SNE visualization showed that the proposed model produces highly discriminative vectors with the Large Margin Cosine Loss model embedded vector being thicker than SoftMax Loss, Semi-Supervised Learning, and a combination of Large Margin Cosine Loss and Semi-Supervised Learning model churn clusters.**

***Index Terms*—Customer Churn, Deep Learning, Large Margin Cosine Loss, Semi-Supervised Learning, Vector Embedding.**

## I. Introduction

Competition in telecommunication is fierce. To proactively foster client loyalty and increase profitability, the operator has to be able to predict customer attrition [1]. Operators attempt to control churn with incoming calls after one or two months of subscription [2].

On the other hand, customers face the problem of deciding which operator is the best as their service provider. It becomes more demanding when companies start offering several types of services, so customers tend to switch between different carriers [3]. Churn occurs when customers unsubscribe from service, leading to revenue losses and operators may need to invest more in advertising and re-offers [4]. Therefore, identifying the causes of churn can help operators to develop strategies to retain their customers [5]. Deep learning is the preferred method for predicting churn in various industries, one of which is the telecommunications industry [6]. One approach is vector embedding [7].

The implementation of the vector embedding system that we mostly encounter today is a recommendation system, especially in the e-commerce industry such as Amazon, Tokopedia, Shopee, and others [8]. Researchers have implemented deep learning for such tasks using SoftMax Loss, which allowed the proposed model to achieve an 81.16% measured by F1 Score for predictive performance. However, SoftMax Loss is inadequate in classification ability because the distance between churn and non-churn is not close, even between two clusters. Despite being known to produce extraordinary classification performance, it is unable to discriminate the embedding vectors of different classes of data [9]. This was the reason researchers apply Large Margin Cosine Loss which is a more advanced and more promising technique using a cosine-distance-based loss function which has been proven in face recognition to increase the discrimination between embedding vectors in different classes [10]. Understanding how mixing unlabeled and labeled input might alter developing algorithms and learning behavior that benefit from this combination are the goals of Semi-Supervised Learning [11]. Meanwhile, a combination of Large Margin Cosine Loss and Semi-Supervised Learning is an interesting thing to apply in this study, providing a performance comparison and experimental evaluation analysis of the vector embedding proposed model.

This study's contribution is to assess and compare the effectiveness of the customer churn vector embedding models with several methods, namely SoftMax Loss, Large Margin Cosine Loss, Semi-Supervised Learning, and a combination of Large Margin Cosine Loss and Semi-Supervised Learning that could be used as a reference for related companies in terms of marketing strategy to determine the tendency of customer churn. Therefore, by using several methods, this study provided a comparison of the performance of vector embedding models whereas previous studies only used SoftMax Loss. This paper is organized as follows: section 2 explains several past studies related to this research, section 3 presents the methodology of the study, section 4 describes the details of the research

This paper was submitted on 24 December 2022. This research was supported by "NVIDIA-BINUS Artificial Intelligence Research & Development Center". The researchers would like to thank those who have provided suggestions and data.

Dinne Ratj is with the Computer Science Department, BINUS Graduate Program – Master of Computer Science Program, Bina Nusantara University, Jakarta, Indonesia (e-mail: dinne.ratj@binus.ac.id)

Tjeng Wawan Cenggoro, Namira Mufida Adien, Ni Putu Putri Ardhia Paramita, and Nabila Putri Sina are with the Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia (e-mail: tjeng.cenggoro@binus.ac.id)

Gregorius Natanael Elwirehardja is with Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia (e-mail: gregorius.e@binus.edu)

Bens Pardamean is with Computer Science Department, BINUS Graduate Program – Master of Computer Science Program and Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia (e-mail: bpardamean@binus.edu)

findings and section 5 provides the conclusion of the study.

## II. Related Work

Many approaches have been developed using deep learning [12]–[14]. Deep learning is the preferred method for churn prediction in telecommunication datasets [6], [15]–[20], gaming [7], [21], [22], banking [23], [24], music streaming [25], [26] and retail industry [27]. Predicting churn behavior with deep learning was originally proposed in 2004 [17]. This research realizes the idea by implementing an autoencoder, deep belief neural network, and multi-layer feedforward network with different configurations. The work was presented on billions of call records from the company's business intelligence systems to predict churn in mobile telecommunications networks. The model produces an accuracy of 77.9% with an operating characteristic curve on the validation data which significantly exceeds random forests with extensive custom feature engineering of 4.7%. The first study was to develop a large-scale mobile game to propose a vector embedding model with Semi-Supervised Learning. This study allows users to act on app recommendations after they quit and choose to play another game [7].

A multilayer perceptron with three layers, each with the same 16 initializer functions and RELU activation, is used for artificial neural network churn prediction. For regularization, a dropout layer is used in the input layer, to prevent overfitting. This study can produce of accuracy more than 80%. Applying deep learning to the vector embedding model has been carried out in our previous study by using SoftMax Loss and can achieve 81.16% of the F1 Score [9]. To continue our research, we propose a vector embedding model with several methods, namely SoftMax Loss, Large Margin Cosine Loss, Semi-Supervised Learning, and a combination of Large Margin Cosine Loss and Semi-Supervised Learning. The use of Large Margin Cosine Loss has been proven to have been applied to face recognition which can increase the discrimination between embedding vectors in different classes [10]. Understanding how mixing unlabeled and labeled input might alter developing algorithms and learning behavior that benefit from this combination are the goals of Semi-Supervised Learning [11]. Meanwhile, a combination of Large Margin Cosine Loss and Semi-Supervised Learning is an interesting thing to be applied in this study which can provide a performance comparison and experimental evaluation analysis of the proposed model so that it can be used as a reference for related companies in terms of marketing to determine the tendency of customer churn. Therefore, by using several methods, this study provided a comparison of the performance of vector embedding models whereas previous studies only used SoftMax Loss.

## III. Materials and methods

### A. DATASET

All the attributes in the dataset are shown in Table I. The dataset telecommunications customer used in this research consists of 3333 rows and 21 columns [28]. Customers' behavior is determined by 19 attributes in the dataset, 4 attributes are categorical, and the rest are numerical. Besides that, we discovered that 1 categorical attribute namely "Account Length" has an allocation that is not discriminative in the definition of customer churn. Therefore, we excluded this characteristic from the vector embedding models. A deep learning model will be trained to predict whether a customer will leave the telecommunications company or not by predicting the churn attribute. Churn is used to describe the loss of consumers who go to competitors. Predicting customer attrition in advance will give businesses highly valuable information to keep and grow their client base [17]. Companies can use probability and income estimation to determine churn intervention and profitability strategies [29].

The dataset for this study is split into three subsets with a ratio of 60:20:20: training, validation, and testing. As discovered from a previous study, loyal and customer churn proportion is imbalanced with a ratio of 15:85. The validation set is used to test the models' performance during training to choose the best models, whereas the training set is used to train the models. A testing dataset is used to compare the effectiveness of the proposed vector embedding model with the best model. A comparison of proposed vector embedding models measured using metrics. The authors quantify the cluster cohesiveness in addition to the F1 Score and Accuracy by computing the distance ratio between the inter-cluster to intra-cluster embedding with cosine similarity for each churn vector and loyal customer embedded vector. Churn is the term used to describe the loss of customers to a business as they switch to rivals.

TABLE I
LISTS OF THE DATASET'S ATTRIBUTES

| Attribute Name | Type | Description |
|---|---|---|
| State | Categorical | State of origin of the phone number's location |
| International plan | Categorical | Whether a client has an international plan or not |
| Voice mail plan | Categorical | Whether a client has a voicemail plan or not |
| Area code | Categorical | An indication of the location from which phone numbers originate |
| Number of voicemail messages | Numerical | how many voicemails customers have left |
| Total day minutes | Numerical | Total daily minutes spent making calls |
| Total day calls | Numerical | All calls made throughout the day |
| Total day charge | Numerical | The total cost of calls made throughout the day |
| Total evening minutes | Numerical | Total minutes used on calls that night |
| Total evening calls | Numerical | Total number of calls made that night |
| Total evening charge | Numerical | The overall cost of the evening's calls |
| Total night minutes | Numerical | Total number of minutes used on calls overnight |
| Total night calls | Numerical | Total number of calls made overnight |
| Total night charge | Numerical | The total cost for late-night calls |
| Total international minutes | Numerical | Minutes used for international calls overall |
| Total international calls | Numerical | Total number of calls made abroad |
| Total international charge | Numerical | All-in cost of international calls |
| Customer service calls | Numerical | The quantity of phone calls to customer service |

### B. EXPERIMENT DETAILS

This section elaborates on the proposed customer churn vector embedding models. The source code used in this research is available at https://github.com/DinneRatj/Vector-Embedding-Model

This research was evaluated with the open-source deep learning framework PyTorch was supported by the NVIDIA-BINUS Artificial Intelligence Research & Development Center. The author trains the proposed model using telecommunication customer datasets. From this dataset, the authors decide how many columns and attributes are needed to train the model. With this, the author will produce an overall system and determine the software flow to the detailed algorithm. Figure 1 visualizes the customer churn vector embedding model architecture. There are four models proposed in this study namely

- Apply deep learning to vector embedding models with SoftMax Loss
- Apply deep learning to embedding vector models with Large Margin Cosine Loss
- Apply deep learning to embedding vector models with Semi-Supervised Learning
- Apply deep learning to embedding vector models with a combination of Large Margin Cosine Loss and Semi-Supervised Learning.

Large Margin Cosine Loss and Semi-Supervised Learning in this study are to continue our research where SoftMax Loss is used in the model proposed from previous studies [9]. The process began by processing the categorical and numerical variables using several layers, the combined intermediate representation is delivered to the following layer. The processing layer is a completely linked layer for the numerical data. The categorical embedding layer, which processes categorical data, is the layer that was employed to create an embedded customer churn vector [9], [30]. The last fully connected layer produces the embedded vector, which is colored orange in Figure 1. The proposed model is trained to classify churn and loyal customers that will generate an embedded vector. Considering loyal customer churn proportion is imbalanced with a ratio of 15:85, Turn out hypothecated that certain treatment is required to adjust the model. This is common knowledge in deep learning research, without any adjustment, The imbalanced dataset will reduce the model's performance [31]–[33]. The churn compared to loyal customer weight ratio was set to 3:7.

The model implemented in this study consists of two input layers, namely 14 numerical data attributes and 4 categorical data attributes. Numerical attributes will be transformed using the Pytorch module with each numerical getting 100 vector representations as output. Meanwhile, categorical attributes will be transformed using the Pytorch module with 51 columns of numerical "state" attributes each of which gets 5-dimensional vector representations as output. Tensors are initialized randomly after embedding layer creation and similarities between categorical will be merged during training. After the input features are transformed, the resulting feature vectors from each layer will be combined or concatenated. After the feature vector is generated by the embedding vector process in the input layer, ReLU is applied to the next layer which contains the resulting feature vector. Due to the existence of a repaired layer in the proposed model, Kaiming initialization is used to initialize the weights and biases, this implementation allows the repaired network to converge and increases the depth of the proposed model [34].

In the Large Margin Cosine Loss, the implementation of the modified loss function is carried out by importing the model Large Margin Cosine Loss. Two numerical attributes get a 120-dimensional vector representation as output. While the value of 0.35 is the optimal value obtained by the inventor [10]. In Semi-Supervised Learning, there is also an encoding and decoding process that is implemented like an Autoencoder. So, there is a layer for the decoding process. After that, the researcher applied the activation function using ReLU with Batch Normalization Layer [35]. The Autoencoder process can reconstruct the output from the received input data. This is an Unsupervised Learning process, where the data used is unlabeled. The model can also classify whether a labeled customer data series determines whether a customer will churn or not, which is a Supervised Learning process. Therefore, in this study, the authors have implemented a model using Semi-Supervised Learning. In a combination of Large Margin Cosine Loss and Semi-Supervised Learning, the coding implementation of the modified loss function is carried out by importing the Large Margin Cosine Loss Model and Semi-Supervised Learning process like the Autoencoder process.

To determine how accurate the model is, the authors use a loss function in the form of Cross Entropy Loss for categorical data, and MSE Loss for numerical data. Cross entropy loss will be added to the weight parameter, due to imbalance data (data that is not churn more than churn, with a ratio of around 85:15). This data imbalance causes the prediction results for the churn class to be not good. The solution is to use Cross Entropy Loss so that the penalty for a churn class is much greater than for a non-churn class. The penalty is enlarged by the inverse ratio of the number of data (15:85). Tuning the model is the stage where the resulting models will be combined for analysis to find out the relationship between data and patterns. The goal is to use previously available data as a reference in the future. The values and types of hyperparameters used follow previous research that will be carried out before and simultaneously with the training process. A standard scaler was used to standardize the numeric data and random permutation to train random data. In addition, Adam with a 0.001 learning rate was used to optimize the proposed model with a configuration from the previous study as the standard [9], [36].

The model was trained with 100 epochs. Model validation is the result of the process of tuning and training the model, so that the model will be used for the model testing process. The results taken from model validation are the best accuracy results or the lowest loss values and will be stored in .pth extension files. The file will then be reloaded for use in the next process, namely the testing model which will be reported as the results of this research. In the model testing phase, the author will determine the success rate of customer churn, performance measurement, and comparison of the performance of the new and old models. The author uses the Confusion Matrix method to determine the accuracy of the method that has been applied to the model. In this study, the author applies a reconfiguration to each proposed model, namely:

- Using the same algorithm on each model proposed
- Using the same dataset in each proposed model
- Using the same hyperparameter values for each proposed model

Replication of this configuration is one of the scopes in this study. The author only provides a comparison of the performance of the proposed models by applying the same configuration replication in each proposed model.

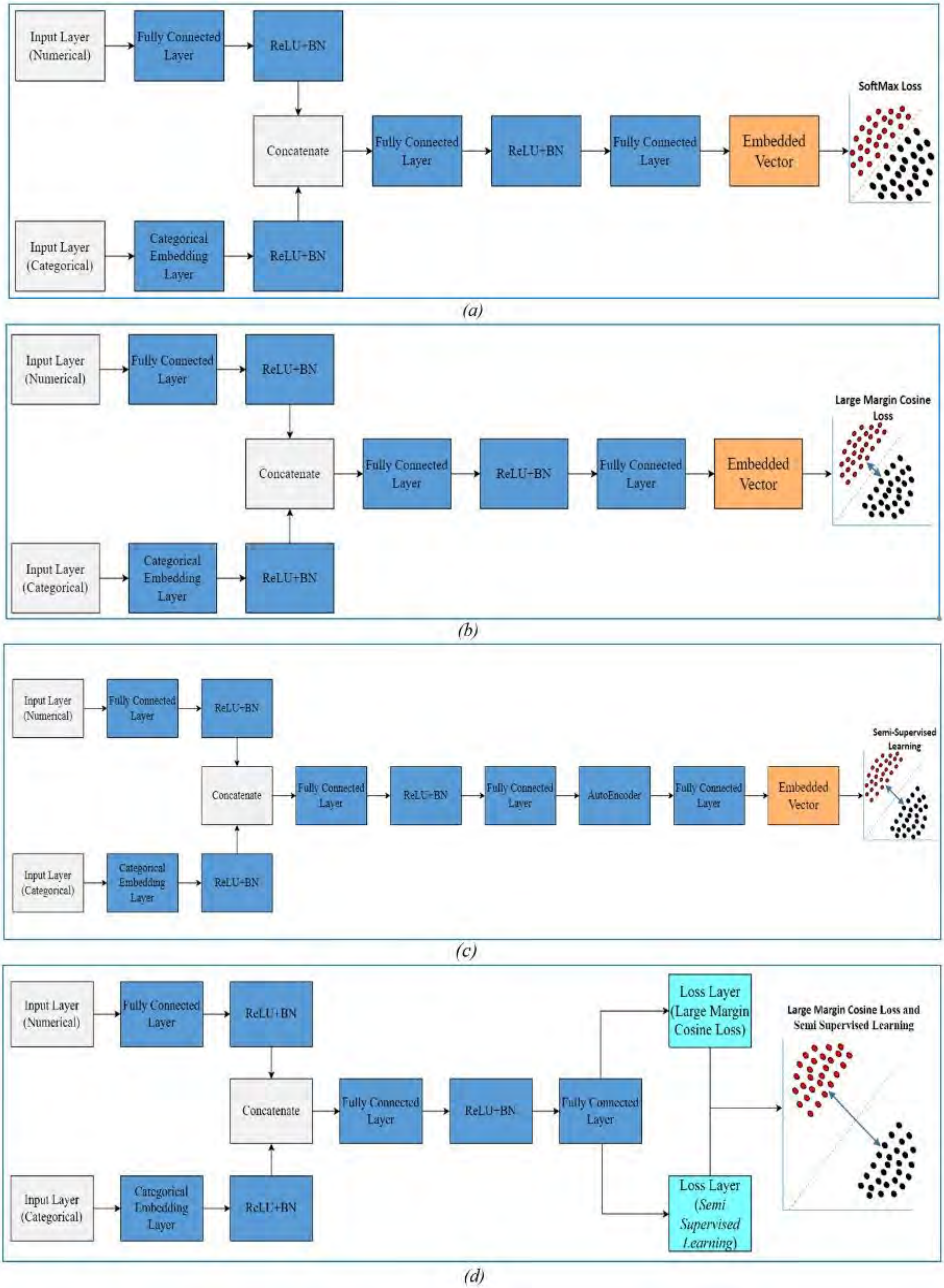


Fig. 1. The Proposed Customer Churn Vector Embedding Model. (a) SoftMax Loss, (b) Large Margin Cosine Loss, (c) Semi-Supervised Learning, and (d) a Combination of Large Margin Cosine Loss and Semi-Supervised Learning.

### C. EVALUATION METHODS

The model loss value and training loss value are displayed together at each epoch. The optimal model is evaluated using a testing dataset to compare the performance model measured by metrics. The best setting of the weight loss ratio of the model is based on the plot. The result of it will be explained in the result section. In addition to the confusion matrix, the authors also measure the cluster cohesiveness formed by each churn vector and loyal customer embedded vector by calculating the distance ratio between intra-cluster and inter-cluster embedding. The calculation of the intra-cluster distance is done by measuring the distance of the features in a cluster with the centre point of the cluster. Calculation of the inter-cluster distance is done by measuring the distance between the features in one cluster and the midpoint in another cluster. This calculation process is done using cosine similarity.

Thus, this calculation can be concluded that a higher distance ratio indicates that the features between clusters have been further separated. The performance of the customer churn vector embedding model is evaluated using the ratio between intra-cluster similarity and inter-cluster similarity, with a higher value ratio. t-SNE [34] is used to measure the model qualitatively to plot the customer churn embedded vector in a 2D space [9]. The measurement of intra-cluster similarity compactness is the cluster cohesiveness created by each churn vector and embedded loyal customers, which is formulated in equation 1. x is the embedded vector and n is the amount of data. The measurement of the similarity from one cluster to another cluster is shown in equation 2. y is the embedded vector from the other cluster and can be data of the loyal class. x is the churn class data or vice versa.

The dataset is split into three subsets, training, validation, and testing, using a ratio of 60:20:20, to determine whether the training model overfits. A training set of 100 epochs is utilized to train the model. The model loss value and training loss value are displayed together at each epoch. Based on the plot, the model's weight loss ratio should be set as low as possible. A testing dataset is used to evaluate the best model and compare its accuracy and F1 Score performance. The measurement of intra-cluster similarity compactness is the cluster cohesiveness created by each churn vector and embedded loyal customers, which is formulated as:

$$\text{Inter} - \text{cluster similarity} = \sum_{i=0}^{n} \frac{Cos(x_i, \bar{y})}{n} \quad (1)$$

$\bar{x}$ is the embedded vector and n is the number of data. The measurement of the similarity from one cluster to another cluster is formulated as follows:

$$\text{Intra} - \text{cluster similarity} = \sum_{i=0}^{n} \frac{Cos(x_i, \bar{x})}{n} \quad 2)$$

$\bar{y}$ is the embedded vector from the other cluster. Data from the churn class may be represented by *x* and data from the loyal class by *y*, or vice versa. With a higher value ratio, the ratio of intra-cluster to inter-cluster similarity is used to evaluate the performance of the model. t-SNE [34] is used to measure the model qualitatively to plot the customer churn embedded vector in a 2D space [9].

## IV. RESULTS & DISCUSSIONS

Figure 2 depicts a comparison of the training loss and validation loss of the customer churn vector embedding model with a ratio of 3:7 of the churn and the loyal weight. At every epoch, the model's validation loss value and training loss value are plotted. The best setting of the weight loss ratio of the model is based on the plot, The optimal model is assessed using a testing dataset to compare the model's performance measured by metrics. In this research, the severity of overfitting was the worst on the model trained using Semi-Supervised Learning, followed by the ones trained using a combination of Large Margin Cosine Loss and Semi-Supervised Learning, SoftMax Loss, and Large Margin Cosine Loss.

The best training loss value is generated by SoftMax Loss, which is approximately 0.15, followed by Semi-Supervised Learning, which generates a value of 0.20, a combination of Large Margin Cosine Loss and Semi-Supervised Learning is around 0.40, and Large Margin Cosine Loss of 0.60. SoftMax Loss also generates the best value for the validation loss, which is around 0.35, followed by Semi-Supervised Learning, which generates a value of 0.60, Large Margin Cosine Loss, which generates a value of 0.90, and a combination of Large Margin Cosine Loss and Semi-Supervised Learning, which generates a value of about 1.0.

When viewed from the difference between the training and validation losses, Large Margin Cosine Loss is the largest, followed by a combination of Large Margin Cosine Loss and Semi-Supervised Learning, Semi-Supervised Learning, and SoftMax Loss. The researcher concludes that Semi-Supervised Learning needs to be analyzed further in further research to determine the best algorithm that can improve the performance of the deep learning model, Additionally, more combinations of Large Margin Cosine Loss and Semi-Supervised Learning.

Figure 3 is a t-SNE visualization of data distribution for loyal customers and customer churn in the proposed model. The green dot represents customer churn, whereas the red dot stands for a loyal client. Principal component 1 is the x-axis, whereas principal component 2 is the y-axis. t-SNE indicated the rendered embedded customer churn vectors of the proposed model have indicated to loyal customers compare churn, the customer churns embedded vectors tend to herd on the top portion of the plot, whereas the loyal cluster is slightly thick. From t-SNE, it shows the customer churn cluster of the Large Margin Cosine Loss model embedded vector is thicker rather than SoftMax Loss, Semi-Supervised Learning, and a combination of Large Margin Cosine Loss and Semi-Supervised Learning model churn clusters on the plots' upper portion. Class churn because of the denser cluster, embedded vectors are separated from the loyal customer cluster. The purpose of the t-SNE visualization is to illustrate that the proposed model produces highly discriminative vectors. The result obtained by observing t-SNE is accurate. The plots for customer churn can be deduced from these customer churn clusters.

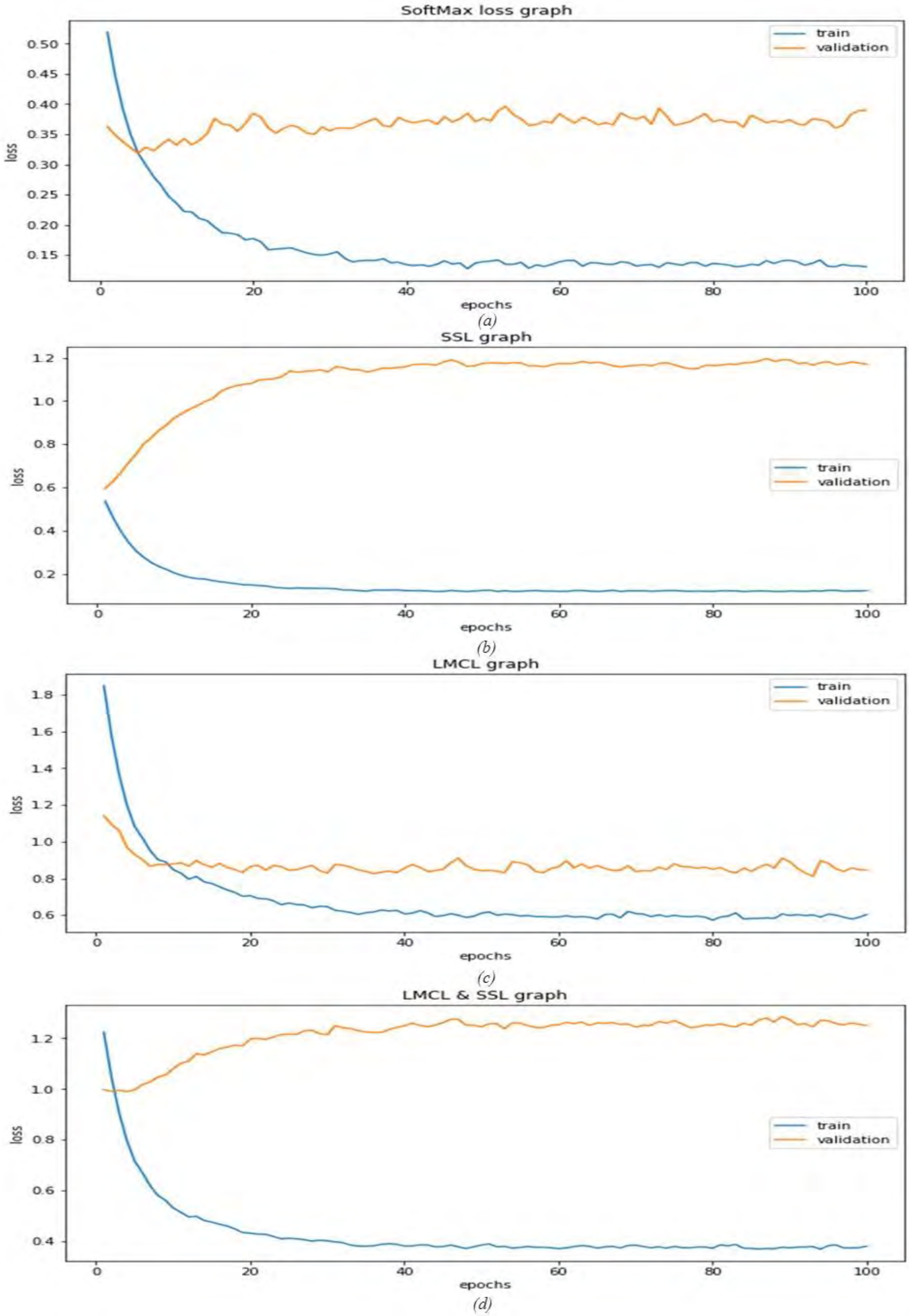


Fig. 2. Train and Validation model loss plot for (a) SoftMax Loss, (b) Large Margin Cosine Loss, (c) Semi-Supervised Learning, and (d) a Combination of Large Margin Cosine Loss and Semi-Supervised Learning.

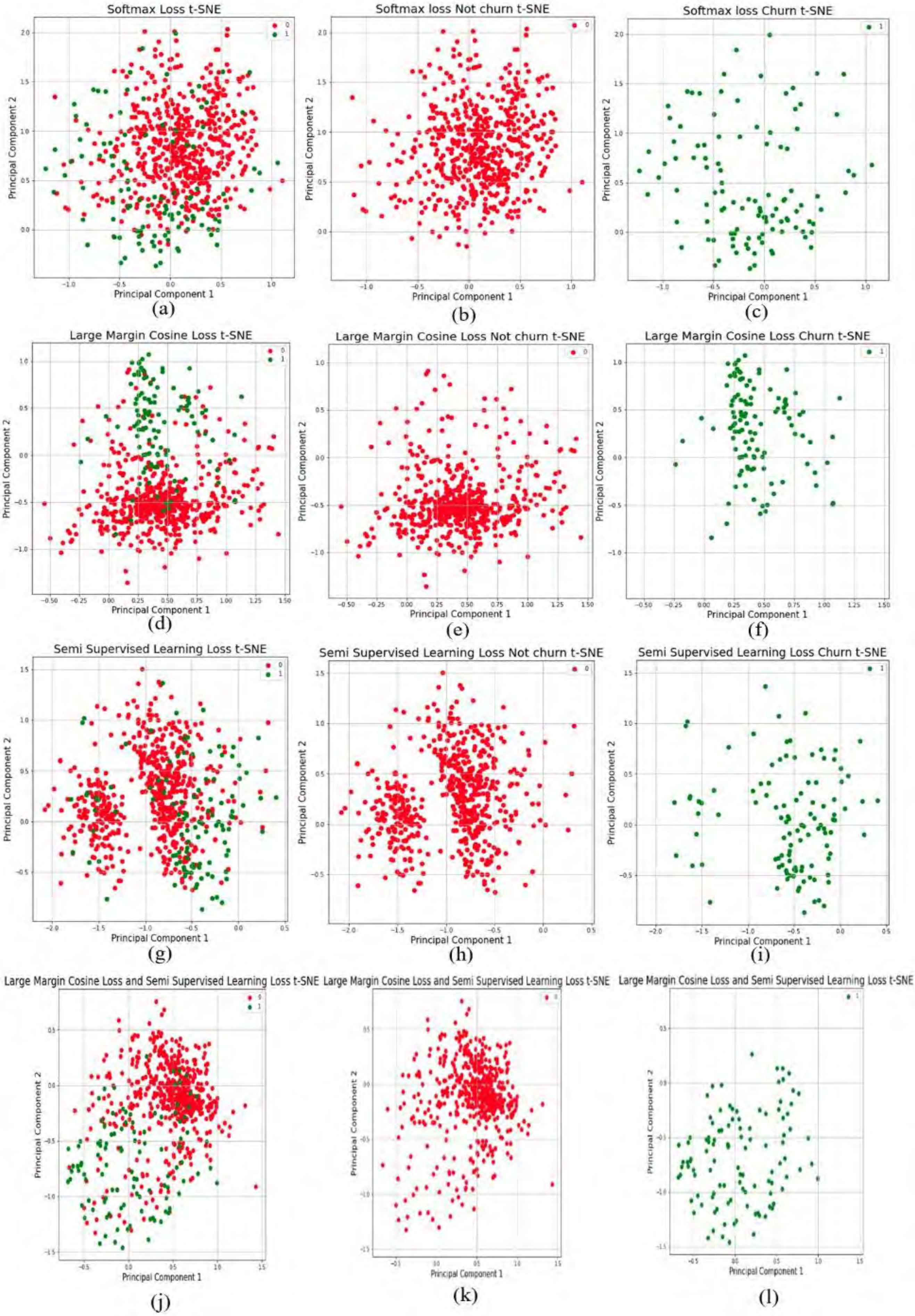


Fig. 3. The t-SNE plot of the embedded vectors generated by: (a) SoftMax Loss, (d) Large Margin Cosine Loss, (g) Semi-Supervised Learning, and (j) a combination of Large Margin Cosine Loss and Semi-Supervised Learning. (c, f, i, and l) The green dot represents churning customers. (b, e, h, and k) The red dot stands for a client that is loyal.

TABLE II
COMPARISON OF PERFORMANCE CUSTOMER CHURN VECTOR EMBEDDING MODEL

| Metrics | SoftMax Loss | Large Margin Cosine Loss | Semi-Supervised Learning | Large Margin Cosine Loss and Semi-Supervised Learning |
|---|---|---|---|---|
| Accuracy | 90.56% | 91.31% | 88.32% | 86.52% |
| F1 Score | 83.04% | 83.74% | 79.37% | 78.63% |
| Loyal Customer inter-class similarity | 0.4825 | 0.3249 | 0.5106 | 0.4227 |
| Loyal Customer intra-class similarity | 0.5884 | 0.5815 | 0.5976 | 0.5844 |
| Loyal Customers intra-class:inter-class similarity ratio | 1.2196 | 1.7895 | 1.1704 | 1.3825 |
| Customers Churn inter-class similarity | 0.5481 | 0.4436 | 0.5479 | 0.5682 |
| Customers Churn intra-class similarity | 0.6492 | 0.6657 | 0.6306 | 0.6865 |
| Customers Churn intra-class:inter-class similarity ratio | 1.1845 | 1.5005 | 1.1509 | 1.2082 |

Table II provides a comparison of the vector embedding models' performance. The proposed model improved from the previous research by 1.88% in terms of SoftMax Loss and the best model is the one trained using the Large Margin Cosine Loss, which allowed it to obtain an 83.73% F1 Score. The results of Loyal Customer intra-class similarity on Semi-Supervised Learning are greater than the other models, namely 0.5976. Semi-Supervised Learning shows the best results on Loyal Customer inter-class similarity, which is 0.5106. For the embedding ratio in Customer Churn intra-class similarity, the expected result also shows a combination of Large Margin Cosine Loss and Semi-Supervised Learning shows a better result with the value of 0.6865. A combination of Large Margin Cosine Loss and Semi-Supervised Learning shows the best result in the Customer Churn inter-class similarity of 0.5682. Meanwhile, in the Loyal Customer intra-class: inter-class similarity ratio, Large Margin Cosine Loss has the best embedding ratio with 1.7895. Large Margin Cosine Loss again has the best ratio with 15005 for customer churn intra-class: inter-class similarity ratio.

In this research, it can be concluded that the proposed models can achieve performance above 78% measured by the F1 Score, and Large Margin Cosine Loss produces the highest performance of 83.74%. However, Semi-Supervised Learning can be analyzed further in future studies to determine the best algorithm that can improve the performance of the model, along with its combination with Large Margin Cosine Loss. The authors apply replicated configurations to each proposed model by using the same algorithm, dataset, and hyperparameter values for each proposed model. This study only provides the comparison performance of the proposed model with the application of the same configuration for each proposed model. In the future, a hyperparameter grid search can be conducted to find the best settings for the vector embedding models.

## V. CONCLUSION

In this paper, a comparison of vector embedding models with several methods was performed, whereas our previous study only used one method. It can be concluded that the four proposed models can achieve performance above 78% measured by F1 Score and Large Margin Cosine Loss produces the highest performance of 83.74%, followed by SoftMax Loss with 83.04%, Semi-Supervised Learning with 79.37% and a combination of Large Margin Cosine Loss and Semi-Supervised Learning with 78.63%. In this research, the severity of overfitting was the worst on the model trained using Semi-Supervised Learning, followed by the ones trained using a combination of Large Margin Cosine Loss and Semi-Supervised Learning, SoftMax Loss, and Large Margin Cosine Loss.

We also depicted the embedded vectors in two-dimensional space and t-SNE to indicate two clusters of customer churn: one cluster can potentially be maintained, and another cluster is probable to quit. The customer churn cluster of the Large Margin Cosine Loss model embedded vector is thicker rather than SoftMax Loss, Semi-Supervised Learning, and a combination of Large Margin Cosine Loss and Semi-Supervised Learning model churn clusters on the plots' upper portion. Additionally, the churn class can be isolated from the loyal customer cluster because of the denser cluster. The Purpose of the t-SNE visualization is to illustrate that the proposed model produces highly discriminative vectors. The result also shows that a combination of Large Margin Cosine Loss and Semi-Supervised Learning shows a greater result in the embedding ratio measured by cosine similarity compared to SoftMax Loss, Large Margin Cosine Loss, and Semi-Supervised Learning.

The researcher concludes that Semi-Supervised Learning needs to be analyzed further in further research to determine the best algorithm that can improve the performance of the deep learning model, Additionally, more combinations of Large Margin Cosine Loss and Semi-Supervised Learning to improve discriminative vectors because in this research, a combination of Large Margin Cosine Loss and Semi-Supervised Learning shows a greater result in the embedding ratio measured by cosine similarity but only can achieve performance 78.63% measured by F1 Score. In future studies, the model can be adjusted to the data owned with the purpose of the model. These methods in this study certainly have their challenges, ranging from datasets and attribute selection to being able to produce a level of similarity in customer behavior to determine whether to churn or not to churn. It is hoped that there will be more research on embedding vectors for customer behavior to deliver superior outcomes, for example, by employing hyperparameter grid search.

REFERENCES


[1] L. Yan, M. Fassino, and P. Baldasare, "Predicting customer behavior via calling links," in Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., pp. 2555–2560, 2005.

[2] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," IEEE Trans Neural Netw, vol. 11, no. 3, pp. 690–696, 2000.

[3] A. A. Khan, S. Jamwal, and M. M. Sepehri, "Applying data mining to customer churn prediction in an internet service provider," Int J Comput Appl, vol. 9, no. 7, pp. 8–14, 2010.

[4] R. Mattison, The telco churn management handbook. Lulu. com, 2006.

[5] C. Kirui, L. Hong, W. Cheruiyot, and H. Kirui, "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining," International Journal of Computer Science Issues (IJCSI), vol. 10, no. 2 Part 1, p. 165, 2013.

[6] O. Adwan, H. Faris, K. Jaradat, O. Harfoushi, and N. Ghatasheh, "Predicting customer churn in telecom industry using multilayer preceptron neural networks: Modeling and analysis," Life Sci J, vol. 11, no. 3, pp. 75–81, 2014.

[7] X. Liu et al., "A semi-supervised and inductive embedding model for churn prediction of large-scale mobile games," in 2018 ieee international conference on data mining (icdm), pp. 277–286, 2018.

[8] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowl Based Syst, vol. 46, pp. 109–132, 2013.

[9] T. W. Cenggoro, R. A. Wirastari, E. Rudianto, M. I. Mohadi, D. Ratj, and B. Pardamean, "Deep learning as a vector embedding model for customer churn," Procedia Comput Sci, vol. 179, pp. 624–631, 2021. doi: 10.1016/j.procs.2021.01.048.

[10] H. Wang et al., "Cosface: Large margin cosine loss for deep face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5265–5274, 2018.

[11] N. Castle, "What is Semi-Supervised Learning," Oracle DataScience. Com, 2018.

[12] N. Dominic, Daniel, T. W. Cenggoro, A. Budiarto, and B. Pardamean, "Transfer learning using inception-resnet-v2 model to the augmented neuroimages data for autism spectrum disorder classification," Communications in Mathematical Biology and Neuroscience, vol. 2021, no. art. 39, 2021. doi: 10.28919/cmbn/5565.

[13] B. Pardamean, H. Soeparno, A. Budiarto, B. Mahesworo, and J. Baurley, "Quantified self-using consumer wearable device: predicting physical and mental health," Healthcare Informatics Research, vol. 26, no. 2, pp. 83–92, 2020. doi: 10.4258/hir.2020.26.2.83.

[14] H. Prabowo, T. W. Cenggoro, A. Budiarto, A. S. Perbangsa, H. H. Muljo, and B. Pardamean, "Utilizing mobile-based deep learning model for managing video in knowledge management system," IJIM, vol. 12, no. 6, pp. 62-73, 2018. doi: 10.3991/ijim.v12i6.8563.

[15] S. Agrawal, A. Das, A. Gaikwad, and S. Dhage, "Customer churn prediction modelling based on behavioural patterns analysis using deep learning," in 2018 International conference on smart computing and electronic enterprise (ICSCEE), pp. 1–6, 2018

[16] S. Cao, W. Liu, Y. Chen, and X. Zhu, "Deep learning based customer churn analysis," in 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6, 2019.

[17] F. Castanedo, G. Valverde, J. Zaratiegui, and A. Vazquez, "Using deep learning to predict customer churn in a mobile telecommunication network," Erişim adresi: https://www. wiseathena. com/pdf/wa_dl. pdf. Erişim tarihi, vol. 30, 2014.

[18] M. Karanovic, M. Popovac, S. Sladojevic, M. Arsenovic, and D. Stefanovic, "Telecommunication services churn prediction-deep learning approach," in 2018 26th Telecommunications Forum (TELFOR), pp. 420–425, 2018.

[19] S. Kumar and M. Kumar, "Predicting customer churn using artificial neural network," in International Conference on Engineering Applications of Neural Networks, pp. 299–306, 2019.

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

[21] E. Lee et al., "Game data mining competition on churn prediction and survival analysis using commercial game log data," IEEE Trans Games, vol. 11, no. 3, pp. 215–226, 2018.

[22] X. Liu et al., "Micro-and macro-level churn analysis of large-scale mobile games," Knowl Inf Syst, vol. 62, no. 4, pp. 1465–1496, 2020.

[23] E. Domingos, B. Ojeme, and O. Daramola, "Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector," Computation, vol. 9, no. 3, p. 34, 2021.

[24] P. S. Patil and N. V Dharwadkar, "Analysis of banking data using machine learning," in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 876–881, 2017.

[25] M. Chen, "A MapReduce-based Artificial Neural Network Churn Prediction for Music Streaming Service," IJCSNS, vol. 22, no. 1, p. 55, 2022.

[26] J. Zhou, J. Yan, L. Yang, M. Wang, and P. Xia, "Customer churn prediction model based on lstm and cnn in music streaming," DEStech Transactions on Engineering and Technology Research, vol. 5, 2019.

[27] A. Dingli, V. Marmara, and N. S. Fournier, "Comparison of deep learning algorithms to predict customer churn within a local retail industry," Int J Mach Learn Comput, vol. 7, no. 5, pp. 128–132, 2017.

[28] "Churn in Telecom's dataset," 2020. https://www.kaggle.com/becksddf/churn-in-telecoms-dataset

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

[30] A. Budiarto, R. Rahutomo, H. N. Putra, T. W. Cenggoro, M. F. Kacamarga, and B. Pardamean, "Unsupervised News Topic Modelling with Doc2Vec and Spherical Clustering," Procedia Comput Sci, vol. 179, pp. 40–46, 2021. doi: 10.1016/j.procs.2020.12.007.

[31] K. Muchtar, F. Rahman, T.W. Cenggoro, A. Budiarto, B. Pardamean, "An Improved Version of Texture-based Foreground Segmentation: Block-based Adaptive Segmenter," Procedia Computer Science, vol. 135, pp. 579–586, 2018. doi: 10.1016/j.procs.2018.08.228.

[32] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," J Big Data, vol. 6, no. 1, pp. 1–54, 2019.

[33] R.E. Caraka, R.C. Chen, H. Yasin, Suhartono, Y. Lee, and B. Pardamean, "Hybrid vector autoregression feedforward neural network with genetic algorithm model for forecasting space-time pollution data," Indonesian Journal of Science and Technology, vol. 6, no. 1, pp. 243–268, 2021. doi: 10.17509/ijost.v6i1.32732.

[34] L. der Maaten and G. Hinton, "Visualizing data using t-SNE.," Journal of machine learning research, vol. 9, no. 11, 2008.

[35] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International conference on machine learning, pp. 448–456, 2015.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

Dinne Ratj received his bachelor's in computer science from Potensi Utama University, Medan, Indonesia in 2011. He is a graduate student of Bina Nusantara University, Jakarta, Indonesia. Currently, He is working as a Cloud Operational Manager at Port Cities. He has over ten years of experience in the information technology industry. He holds several certifications in information technology.

Tjeng Wawan Cenggoro is an AI researcher whose focus is on the development of Deep Learning algorithms. He has led several research projects that utilize Deep Learning for Computer Vision, which is applied to various cases such as indoor video analytics and plant phenotyping. He has published over 50 peer-reviewed publications and reviewed for prestigious journals such as Scientific Reports and IEEE Access. He also holds 2 copyrights for AI-based video analytics software.

Namira Mufida Adien is a researcher at Bina Nusantara University. She received her undergraduate degree in computer science from Bina Nusantara University.

Ni Putu Putri Ardhia Paramita is a researcher at Bina Nusantara University. She received her undergraduate degree in computer science from Bina Nusantara University.

Nabila Putri Sina is a researcher at Bina Nusantara University. She received her undergraduate degree in computer science from Bina Nusantara University.

Gregorius Natanael Elwirehardja received his bachelor's and master's degree in computer science from Bina Nusantara University, Jakarta. He began his career in research as a research assistant in Bioinformatics & Data Science Research Center of Bina Nusantara University, as well as NVIDIA-BINUS Artificial Intelligence Research & Development Center. His research interests include applications of machine learning in various fields, such as computer vision, agriculture, psychology, medical, and genomics. He is also a certified instructor at NVIDIA Deep Learning Institute.

Bens Pardamean has over thirty years of global experience in information technology, bioinformatics, and education. His professional experience includes being a practitioner, researcher, consultant, entrepreneur, and lecturer. He currently holds a dual appointment as Director of Bioinformatics & Data Science Research Center (BDSRC) | AI Research & Development Center (AIRDC) and Professor of Computer Science at Bina Nusantara (BINUS) University in Jakarta, Indonesia. He earned a doctoral degree in informatics research from University of Southern California (USC), as well as a master's degree in computer education and a bachelor's degree in computer science from California State University at Los Angeles (USA).