Deep Neural Networks: Predictive Research on Customer Turnover Caused by Enterprise Marketing Problems

Ning Li*, Lihan Gu
Business School, Taishan University, Tai'an, Shandong, 271000, China
Corresponding address: No. 525, Dongyue Street, Daiyue District, Tai'an City, Shandong 271000, China
Email: lning1979@outlook.com

Keywords: deep neural network, enterprise marketing, classification model, customer turnover

Received: April 12, 2024

Customer turnover prediction can assist enterprises in identifying potential lost customers early and formulating marketing strategies to retain them. This paper used telecom enterprise A as an illustrative example for customer turnover prediction. A balanced dataset was obtained through the synthetic minority oversampling technique (SMOTE) algorithm. Feature selection was conducted using the IV value. Additionally, the Inception v1 structure was optimized based on a deep neural network to design a deep convolutional neural network (CNN). Experiments were performed on the dataset of telecom enterprise A and the customer turnover datasets from Kaggle. On the Kaggle datasets, the deep CNN demonstrated superior classification performance compared to conventional approaches such as random forest (RF) and XGBoost. It exhibited a higher recall rate, F_2 score, and area under the curve (AUC) value. The dataset of telecom enterprise A enhanced the prediction effectiveness of the deep CNN after processing by the SMOTE algorithm, and a recall rate of 0.97, a F_2 score of 0.98, and an AUC value of 0.98 were achieved. These results show the reliability of the deep CNN for customer turnover prediction and its practical applicability.

Povzetek: Članek analizira napovedovanje odhoda strank z uporabo optimizirane globoke nevronske mreže, ki doseže boljše rezultate kot tradicionalni pristopi, s poudarkom na povečani zanesljivosti napovedi.

1 Introduction

Under the influence of economic development and heightened market competition, enterprises face unprecedented challenges to their survival and development. In this context, customer loss has become a closely monitored issue across various industries. For customer-centric enterprises, customer resources directly impact the survival of the business. The loss of customers signifies a decline in market share, and enterprises need to invest a significant amount of resources to attract new customers. Furthermore, a significant customer loss may lead to negative word-of-mouth dissemination, adversely affecting the long-term development of the enterprise. Customer turnover prediction plays a crucial role in enabling enterprises to proactively implement marketing strategies to retain customers before they turnover [1], which holds significant value for enterprise development [2] and has been extensively researched [3]. The related works are summarized in Table 1.

Table 1: A summary table of related works.

	Method	Result
Zhou et al. [4]	Logistic regression	The effectiveness of the model was validated through analysis of survey data.
Maw et al. [5]	Data sampling techniques and six classifiers	The random forest (RF) classifier exhibited good classification performance.
Tavassoli and Koosha [6]	Three integrated bagging and boosting-based classifiers	The hybrid method showed better accuracy and precision for customer churn prediction.
Zhu et al. [7]	Long short-term memory (LSTM)	Its performance was better than the baseline methods.

Based on current research, there have been many methods studied in customer turnover prediction. However, the majority of them use machine learning methods and give less consideration to deep learning methods. There is still potential for further improvement in the accuracy of customer turnover prediction. Deep neural networks (DNN) are algorithms with multiple layers of nodes that can achieve better results in many tasks. As customer turnover prediction is a binary classification problem, it can also be addressed using DNN. This article investigated the usability of DNN in customer turnover prediction and validated its effectiveness through experiments on a dataset. A telecom company was taken as an example to offer marketing suggestions. The research provides reliable references for enterprise customer management and marketing strategies, contributing to maintaining competitiveness. Additionally, it provides theoretical references for the further application of DNN and other methods in this field.

2 Customer turnover prediction

2.1 Customer turnover and causes

The reasons for customer turnover can typically be categorized into two types:

(1) natural turnover due to customer's move, change of job, etc.;

(2) unnatural turnover due to reasons such as poor service and poor marketing by enterprises.

Typically, the cost of attracting new customers is higher than the cost of maintaining old ones, and this is because:

(1) customers trust enterprises more;

(2) customers require less marketing costs;

(3) customers are familiar with the company's products and services and have a higher willingness to spend.

Therefore, the prediction of customer turnover has become an urgent need for enterprises [8] and can help enterprises identify customers showing early signs of turnover, allowing them to implement suitable methods for retention. Additionally, it facilitates a deeper understanding of customer needs through the analysis of customer data, which helps enterprises adjust their marketing and service strategies to foster the development of new customers while maintaining the stability of existing ones.

Customer turnover prediction is a dichotomous problem [9], which is carried out through the steps shown in Figure 1.

Figure 1: Customer turnover forecast.

Firstly, the process begins with the collection of the original customer data. Subsequently, a certain level of processing is applied to enhance its reliability. Following this, relevant indicators associated with customer turnover are selected as features from the processed data. These features are then input into the chosen classification model for training, which may include support vector machines (SVM) and neural networks (NN) [10]. After the completion of the training, the model can be utilized to forecast customer turnover or retention. The results obtained from the prediction are then employed to develop targeted marketing strategies to retain customers and mitigate turnover.

2.2 Customer turnover analysis for telecom company A

Telecom enterprise A in Shandong was taken as an example for analysis. The implementation of number portability has strengthened the mobility of telecom customers, leading to a significant increase in the rate of customers leaving the network of major telecom enterprises. Due to market saturation, the scope for new subscriber growth has diminished, intensifying competition among telecom enterprises. Consequently, mobile virtual network operators are increasingly focusing on improving customer retention and reducing off-network rates through effective marketing strategies. For telecom enterprise A, several reasons contribute to its customer turnover:

(1) Customer's personal reasons: some customers tend to choose companies offering lower prices or a broader range of services;

(2) competitive enterprise attraction: competitive firms introduce more attractive products;

(3) factors related to the enterprise: the quality of products or services is poor, and charges are excessively high.

Drawing on actual business experience, this paper conducted a customer turnover analysis for enterprise A. A subset of customer data from August to October 2023 was randomly selected from the database. The distribution of valid data, obtained after eliminating abnormal and incomplete data, is presented in Table 2.

Table 2: Original data set.

Sample size	Number of customers in the network	Number of customers lost
7,909	7,598 (96.07%)	311 (3.93%)

The data collection phase revealed that the number of churned customers was much smaller than the number of non-churned (in the network) customers, i.e., the dataset was unbalanced, which will have an impact on the subsequent prediction results. Therefore, SMOTE [11] was used to process the original dataset, and the process is shown below.

(1) Based on the K-nearest neighbor algorithm, K nearest neighbors of the sample class with a small proportion were calculated.

(2) K samples were randomly selected for random linear interpolation.

(3) New data x_{new} was established for the sample class with a small proportion:

$x_{new} = x_i + rand(0,1) \times (x_j - x_i)$,

where x_i refers to a sample in the sample class with a small proportion and x_j is a sample randomly selected from K nearest neighbors.

(4) The old and new data were merged to get a balanced dataset.

The new dataset obtained after SMOTE processing is presented in Table 3.

Table 3: Comparison of old and new datasets.

	Sample size	Number of customers in the network	Number of customers lost
Original dataset	7,909	7,598 (96.07%)	311 (3.93%)
Balanced dataset	15,196	7,598 (50%)	7,598 (50%)

Based on actual business experience, the following indicators that may be related to customer turnover were selected for the subsequent prediction of the customer turnover of enterprise A. See Table 4.

Table 4: Customer turnover characteristics.

Serial number	Indicator	Serial number	Indicator
1	Customer number	13	Overflowing call minutes
2	Gender	14	Call duration within the plan
3	Age	15	Call duration outside the plan
4	Length of time in the network	16	Average monthly mobile data
5	Number of secondary cards	17	Overflowing mobile data
6	Whether integrated (for broadband or ITV)?	18	Mobile data within the plan
7	Have the customer signed the contract?	19	Mobile data outside the plan
8	Whether 4G network coverage?	20	Cumulative number of complaints per year
9	Whether or not the bank card is linked?	21	Cumulative number of fault declarations per year
10	Average monthly minutes of phone calls	22	Number of calls to customer service from other networks in the last three months
11	Average number of outgoing calls per month	23	Number of months in arrears
12	Average monthly calling duration	24	Number of months remaining until credit activity expires

The large number of features in Table 3 may lead to overfitting of the model; therefore, feature selection was needed to retain the more important features. The information value (IV)-based method was selected [12].

Before calculating the IV, the weight of evidence (WOE) value was calculated. The WOE value for the i-th group is as follows:

$$WOE_i = \ln\left(\frac{\rho_{y_i}}{\rho_{n_i}}\right) = \ln\left(\frac{y_i}{y_T}\right) - \ln\left(\frac{n_i}{n_T}\right),$$

where y_i is the quantity of lost customers in the i-th group of features, y_T is the total quantity of lost customers, n_i is the quantity of non-lost customers in the i-th group of features, n_T is the total quantity of non-lost customers, ρ_{y_i} is the proportion of the lost customers in the i-th group of features, and ρ_{n_i} is the proportion of the non-lost customers in the i-th group of features.

The IV of the i-th group of features is:

$$IV_i = \left(\rho_{y_i} - \rho_{n_i}\right) * WOE_i = \left(\frac{y_i}{y_T} - \frac{n_i}{n_T}\right) * \ln\left(\frac{y_i}{y_T}\right) - \ln\left(\frac{n_i}{n_T}\right).$$

The larger the IV value, the more distinct the distinction between lost and non-lost customers was after feature grouping, i.e., the feature had a stronger predictive capacity. The features were categorized according to the magnitude of the IV, as shown in Table 5.

Table 5: IV and predictive capabilities.

IV	Forecasting capability
< 0.02	None
0.02-0.1	Weak
0.1-0.3	Moderate
0.3-0.5	Relatively strong
> 0.5	Strong

Only the features with IV > 0.1 were retained in the prediction, and the ten features obtained after screening are shown in Table 6.

Table 6: Features after screening.

Serial number	Indicator	Serial number	Indicator
1	Whether integrated?	6	Average number of outgoing calls per month
2	Number of secondary cards	7	Months in arrears
3	Average monthly minutes of phone calls	8	Cumulative number of complaints per year
4	Call duration within the plan	9	Mobile data outside the plan

5	Number of calls to customer service staffs from other networks in the last three months	10	Average monthly mobile data

In Table 6, the features "whether integrated" and "number of secondary cards" can reflect the value of customers. Generally, customers who choose integration services and have a higher number of secondary cards tend to have higher values and lower possibilities of turnover. The features "average monthly minutes of phone calls," "call duration within the plan," "average number of outgoing calls per month," "mobile data outside the plan," and "average monthly mobile data' can reflect customers' communication behaviors. If customers' communication behavior decreases, the possibility of turnover also increases. The features "number of calls to customer service staffs from other networks in the last three months," "months in arrears," and "cumulative number of complaints per year" can reflect customers' business behavior to some extent, indicating their satisfaction level with the service. If customers have more calls with customer service staffs from other networks, more arrears, or more complaints, then there is a higher possibility of turnover.

3 Prediction method based on a deep neural network

DNN includes multiple hidden layers, enabling it to learn intricate feature representations. It has demonstrated success in various applications, such as image processing and speech recognition [13]. In comparison, CNN is a specialized type of DNN. Unlike traditional DNNs, CNNs exhibit superior performance in capturing complex features. Hence, this paper designed a customer turnover prediction model leveraging the capabilities of CNN.

CNN demonstrates excellent performance in processing images, signals, text, and other data types [14]. Its effectiveness can be enhanced by increasing the depth or width of the network. However, this approach often results in a substantial increase in network parameters. Google's open-source deep CNN, Inception, addresses this challenge through a network-broadening technique involving multi-scale operations [15]. This strategy allows for improving network performance while simplifying the overall network structure. In the context of customer turnover prediction, this paper incorporated the Inception structure into CNN.

In Inception v1, multiple convolutional kernels are employed to extract features, and 1×1 convolutions are used to reduce the feature mapping hierarchy, thereby minimizing network parameters. To enhance feature extraction performance, this paper introduced modifications to the Inception v1 structure. The architecture of the deep CNN classification model designed for customer turnover prediction is illustrated in Figure 2.

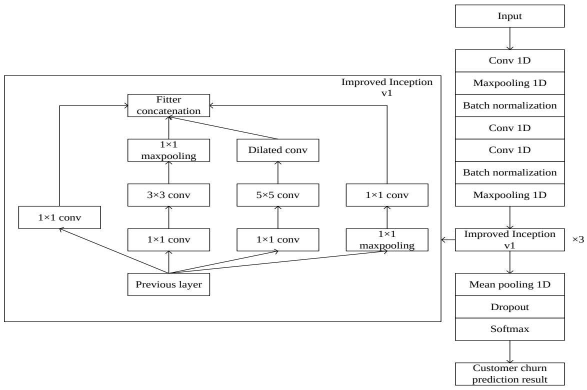

Figure 2: Deep CNN-based customer turnover prediction model.

As shown in Figure 2, the improved Inception v1 module adds pooling and dilated convolution layers after the two groups of parallel convolution layers in the middle group to further enhance the selection of important features and reduce attention on unnecessary information. In the overall CNN model, features are extracted through convolutional pooling and a batch normalization layer is added to enhance the model's generalization ability. One-dimensional convolution used the rectified linear unit (ReLU) activation function, and its convolution operation is written as:

$$x_j^l = ReLU\left(\sum_i w_j^l * x_i^{l-1} + b_j^l\right),$$
$$ReLU(x) = max(0, x),$$

where x_j^l is the j-th output of the l-th layer, w_j^l is the j-th convolution kernel of the l-th layer, and b_j^l is the j-th bias of the l-th layer.

The pooling layer achieved feature dimensionality reduction by downsampling:

(1) maximum pooling: the maximum value among all pixel points within the sub-block was taken as the result (Figure 3);

(2) mean pooling: the average value of all pixel points within the sub-block was taken as the result (Figure 4).

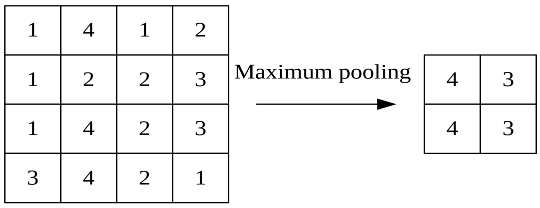

Figure 3: Maximum pooling

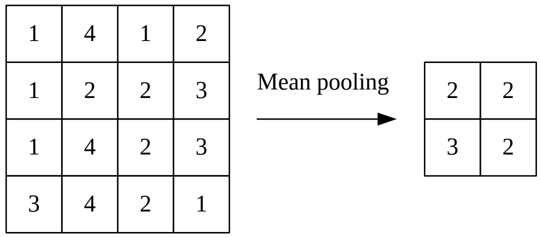

Figure 4: Mean pooling

The purpose of dilation convolution [16] is to increase the receptive field while maintaining the same parameters, which can be written as:

$$r_n = r_{n-1} + (k' - 1)\prod_{i=1}^{n-1} s_i,$$

where r_n is the receptive field of the current layer, r_{n-1} is the receptive field of the last layer, k' is the convolution kernel of dilated convolution, $k' = k + (k-1)(r-1)$ (k is the standard convolution kernel and r is the dilation rate), and s_i is the convolution step length of the i-th layer.

While increasing the receptive field, the dilation convolution will not affect the feature map size, thus it can optimize the training effect of the model while obtaining more information. In this paper, the dilation rate = 2. An example is shown in Figure 5.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 5 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 7 & 0 & 8 & 0 & 9 \end{bmatrix}$$

Figure 5: The dilated convolution when the dilation rate = 2

Eventually, the high-order features extracted from the deep CNN were classified in the softmax layer, the output was converted to the classification probability, and then the category with the highest probability was used as the result to realize the customer turnover prediction. As a classification model, the deep CNN was trained with a binary cross-entropy loss function, written as:

$$loss = -\frac{1}{n}\sum_{i=1}^{n} y_i \log y_i' + (1 - y_i)\log(1 - y_i'),$$

where y_i is the real category, y_i' is the predicted category, and n is the number of training samples.

4 Results and analysis

4.1 Experimental setup

A model was constructed using Keras, with Python as the programming language. In the deep CNN, the optimization algorithm employed was Adam. The parameters of the model were determined through multiple times of experiment (Table 7).

Table 7: The parameter setting of the deep CNN

Parameter	Value
The size of the convolution kernel	(2,3,4)
The number of convolution kernels	32
Learning rate	0.001
Batch size	128
Dropout rate	0.5
Maximum number of iteration	100

In addition to the customer turnover dataset obtained from enterprise A, four additional customer turnover datasets were selected from Kaggle [17] for experimental purposes. The data distribution is presented in Table 9.

Table 9: Kaggle customer turnover dataset.

Dataset	Sample size	Number of attributes	Percentage of lost customers
Telecom-1	100,000	100	49.56%
Insurance	33,908	17	11.70%
BankChurners	10,127	23	16.07%
Customertravel	954	7	23.48

All datasets were processed by SMOTE balance and IV-based feature screening. A ten-fold cross-test was performed, and the final results were averaged. The evaluation of the model was based on a confusion matrix (Table 8).

Table 8: Confusion matrix.

		Forecasted category	
		Positive category	Negative category
Real category	Positive category	TP	FN
	Negative category	FP	TN

Specific indicators included:

(1) Accuracy:
$A = \frac{TP+TN}{TP+FN+FP+TN}$.

(2) Precision:
$P = \frac{TP}{TP+FP}$.

(3) Recall rate:
$R = \frac{TP}{TP+FN}$.

(4) F_β score:
$F_\beta\ score = (1+\beta^2)\cdot\frac{P\cdot R}{\beta^2\cdot P+R}$.

In customer turnover prediction, more emphasis should be placed on predicting potential turnover, i.e., more emphasis should be placed on recall rate R, so in this paper, $\beta = 2$. The F_2 score was taken as the indicator during model evaluation.

(5) area under the curve (AUC): the area under the receiver operator characteristic curve composed of false positive rate (FPR) and true positive rate (TPR), which can describe the advantages and disadvantages of the classification model. The closer the value is to 1, the better the performance is. The FPR and TPR are calculated:

$FPR = \frac{FP}{TN+FP}$,

$TPR = \frac{TP}{TP+FN}$.

4.2 Experimental results

First, on the Kaggle dataset, to verify the performance of the proposed method, it was compared with the following approaches:

(1) RF [18],

(2) gradient-boosted decision tree (GBDT) [19],

(3) extreme gradient boosting (XGBoost) [20],

(4) back-propagation neural network (BPNN) [21],

(5) CNN.

The results obtained after using different datasets are shown in Table 10.

Table 10: Prediction results obtained using the Kaggle dataset (bold indicates optimal values).

		A	P	R	F_2 score	*AUC*
Telecom-1	RF	0.62	0.61	0.61	0.61	0.58
	GBDT	0.64	0.62	0.66	0.65	0.59
	XGBoost	0.63	0.62	0.63	0.63	0.61
	BPNN	0.64	0.63	0.63	0.63	0.63
	CNN	0.65	**0.64**	0.68	0.67	0.66
	Deep CNN	**0.66**	**0.64**	**0.71**	**0.69**	**0.71**
Insurance	RF	0.91	0.66	0.41	0.44	0.61
	GBDT	0.91	0.66	0.42	0.45	0.63
	XGBoost	0.91	0.63	0.47	0.50	0.64
	BPNN	0.91	0.64	0.55	0.57	0.66
	CNN	0.91	**0.63**	0.58	0.59	0.68
	Deep CNN	**0.92**	**0.63**	**0.61**	**0.61**	**0.71**
BankChurners	RF	0.95	0.91	0.77	0.79	0.77
	GBDT	0.96	0.92	0.83	0.85	0.82
	XGBoost	**0.97**	0.92	0.88	0.89	0.89
	BPNN	**0.97**	0.92	0.89	0.90	0.91
	CNN	**0.97**	0.92	0.91	0.91	0.92
	Deep CNN	**0.97**	**0.93**	**0.92**	**0.92**	**0.93**
Customertravel	RF	0.87	0.74	0.66	0.67	0.71
	GBDT	**0.88**	0.77	0.68	0.70	0.73
	XGBoost	0.87	0.74	0.66	0.67	0.75
	BPNN	0.87	0.75	0.68	0.69	0.76
	CNN	**0.88**	0.75	0.73	0.73	0.77
	Deep CNN	**0.88**	**0.76**	**0.76**	**0.76**	**0.78**

From Table 8, it is evident that all the results obtained by the deep CNN model designed in this paper surpassed those of the other algorithms on the four datasets. In comparison, machine learning methods such as RF, GBDT, and XGBoost demonstrated average performance, and the DNN method notably outperformed the traditional shallow neural network, BPNN. This result demonstrated the advantage of DNN in feature extraction. Although methods like XGBoost achieved good accuracy on datasets such as BankChurners and Customertravel, their performance was suboptimal in terms of recall rate. In contrast, the deep CNN designed in this paper exhibited a substantial improvement in recall rate compared to RF and other machine learning methods, highlighting its effectiveness in predicting customer turnover.

The comparison between the deep CNN and CNN showed that the former further optimized the network structure. The recall rate, F_2 score, and AUC value were 0.71, 0.69, and 0.71, respectively, when using the Telecom-1 dataset, which showed improvements of 0.03, 0.02 and 0.05, respectively, compared to the CNN. The recall rate, F_2 score, and AUC value were 0.61, 0.61, and 0.71 respectively when using the Insurance dataset, showing an improvement of 0.03, 0.02, and 0.03, respectively, compared to the CNN. When utilizing the BankChurners dataset, the recall rate, F_2 score, and AUC value were found to be 0.92, 0.92, and 0.93, respectively, which showed an increase of 0.01, 0.01, and 0.01, respectively. The recall rate, F_2 score, and AUC value were 0.76, 0.76, and 0.78 when using the Customertravel dataset, indicating an improvement of 0.03, 0.03, and 0.01, respectively. These results demonstrated the performance of the deep CNN in forecasting customer turnover.

Next, on the actual customer dataset from enterprise A, the performance of the designed prediction method was evaluated using an ablation experiment, and the results are presented in Table 11.

Table 11: Ablation experiment.

	A	P	R	F_2 score	AUC
The deep CNN method	0.98	0.99	0.97	0.98	0.98
Remove SMOTE	0.94(-0.04)	0.99 (-	0.94 (-	0.95 (-	0.96 (-

preprocessing		0.00)	0.03)	0.03)	0.02)
Remove feature selection	0.77 (-0.21)	0.75 (-0.24)	0.76 (-0.21)	0.76 (-0.22)	0.73 (-0.25)
Remove the improved Inception v1 structure	0.89 (-0.09)	0.91 (-0.08)	0.90 (-0.07)	0.90 (-0.08)	0.91 (-0.07)

From Table 11, it can be observed that the impact of SMOTE preprocessing on model prediction results was relatively small. Removing SMOTE preprocessing led to a decrease in the model's F_2 score by 0.03 and a decrease in AUC value by 0.02. The most significant factor affecting customer turnover prediction performance was feature selection. Without proper feature selection, the excessive complexity of data resulted in insufficient training of the model and inaccurate predictions. In the absence of a feature selection module, the model's F_2 score decreased by 0.22 to 0.76, and the AUC value decreased by 0.25 to 0.73, highlighting the importance of feature selection. Removing the improved Inception v1 structure resulted in a decrease in the model's F_2 score by 0.08 and a decrease in AUC value by 0.07, indicating that the improved Inception v1 structure had a positive effect on enhancing predictive accuracy. The structure could enhance the accuracy of customer turnover prediction by obtaining more feature information.

The customer turnover prediction results of the CNN and deep CNN were compared using a balanced dataset (Tables 12 and 13).

Table 12: CNN confusion matrix.

		Forecasted category	
		Customers in the network	Lost customers
Real category	Customers in the network	7,311	287
	Lost customers	265	7,333

Table 13: Deep CNN confusion matrix.

		Forecasted category	
		Customers in the network	Lost customers
Real category	Customers in the network	7,401	197
	Lost customers	99	7,499

Comparing Tables 12 and 13, it can be found that the deep CNN showed better results in the prediction of lost customers. After calculation, the R value of the CNN was 0.96, the F_2 score was 0.96, and the AUC value was 0.96, while the R value of the deep CNN was 0.97, the F_2 score was 0.98, and the AUC value was 0.98, which verified that the deep CNN had more stable classification effect on the customer dataset of enterprise A and could realize accurate prediction of lost customers.

5 Discussion

Customer turnover prediction is a highly complex and significant problem. Given the limited application of deep learning methods in this field, this study primarily focused on investigating deep DNN, proposed a novel deep CNN model, and validated its reliability through experiments conducted on two datasets.

Compared with the current discussions on the availability of proposed methods only on public datasets or practical datasets, this paper validated the model's applicability on both types of datasets to understand its adaptability to different data. The results showed that on the Kaggle dataset, the deep CNN demonstrated significant advantages compared to other classification methods, such as RF and GBDT. It achieved good prediction results regardless of which dataset was applied. Furthermore, compared to ordinary CNNs, deep CNNs enhanced feature extraction capability by deepening the network structure, further improving accuracy in predicting customer turnover.

Then, through the analysis of ablation experiments and confusion matrix on the actual customer turnover dataset of telecom company A, it can be observed that the proposed improvements were beneficial for improving the effectiveness of customer turnover prediction. SMOTE preprocessing, feature selection, and adding an improved Inception v1 structure all contributed to enhancing model performance, and feature selection played the most significant role. From a comparison of confusion matrices, it is evident that deep CNN performed better in distinguishing between users in the network and churned customers, making it more suitable for practical customer turnover prediction in telecom company A.

Aiming at the current customer turnover situation of telecom enterprise A, based on the realization of customer turnover prediction using a deep CNN, this paper puts forward some suggestions on marketing strategies to retain the customers that may be lost, as follows.

(1) A dedicated individual responsible for customer retention is crucial. This person plays a pivotal role in identifying potential lost customers early through processing and analyzing customer data. They need to formulate a comprehensive list of target lost customers and establish clear goals and tasks for effective customer retention.

(2) Use a personalized marketing strategy

① For customers with saturated package expenses, marketing activities such as upgrading, downgrading, and horizontal transferring are provided according to the customers' actual usage, or customized packages are provided according to customers' habits, to enhance customer satisfaction.

② For customers with a high number of complaints, the company can attract customers to renew their business by providing preferential activities, giving away call duration/mobile data, and other services, promptly solving the problems that customers encounter in the process of using the product, and maintaining regular communication with customers to calm them down.

(3) The company should strengthen business integration, providing customers with more products and services. It should increase business integration efforts based on understanding customers' actual usage needs, integrate services such as "broadband + TV + secondary card" in addition to the traditional voice communication and mobile data products, and strengthen the integrated marketing of the family business and government-enterprise business to improve customer loyalty.

(4) The company should upgrade its products, accelerate the research and development of digital products, build a smart home business, consolidate and promote the volume of home business, and improve the competitiveness of products.

(5) The company should strengthen external cooperation, break down industry barriers, strengthen cooperation with other audio and video software and network platforms, and implement cross-industry convergence products to provide customers with more choices, enhance customer stickiness, and reduce the loss of customers.

6 Conclusion

In this paper, a customer turnover prediction model based on a deep CNN was designed. Through experiments conducted on both real-world telecom enterprise A data and the Kaggle dataset, the developed method demonstrated superior results in customer prediction compared to conventional methods such as RF and XGBoost. The deep CNN particularly excelled in achieving a higher recall rate, F_2 score, and AUC value compared to the conventional CNN. Feature selection played a significant role. The results on different datasets all proved that the proposed deep CNN exhibited excellent discriminative ability, enabling accurate prediction of customer turnover. It can be further promoted and applied in practice, aiming to provide reliable support for enterprise marketing. However, this study also has some limitations. For example, the dataset collected from actual telecom company A was relatively small and did not consider a more comprehensive range of customer characteristics. In future work, analysis of customer data from actual telecom companies will be conducted to analyze more customer features and collect datasets from different telecom companies to validate the effectiveness of the proposed method.

References

[1] Khalid L F, Mohsin Abdulazeez A, Zeebaree D Q, Ahmed F Y H, Zebari D A (2021) Customer churn prediction in telecommunications industry based on data mining. *2021 IEEE Symposium on Industrial Electronics & Applications (ISIEA), Langkawi Island, Malaysia*, pp. 1-6. https://doi.org/10.1109/ISIEA51897.2021.9509988.

[2] VLN R K, Deeplakshmi P (2021) Dynamic churn prediction using machine learning algorithms - Predict your customer through customer behaviour. *2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India*, pp. 1-6. https://doi.org/10.1109/ICCCI50826.2021.9402369.

[3] Alhaqui F, Elkhechafi M, Elkhadimi A (2022) Machine learning for telecoms: From churn prediction to customer relationship management. *2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), Soyapango, El Salvador*, pp. 1-5. https://doi.org/10.1109/ICMLANT56191.2022.9996496.

[4] Zhou Y, Gong F, Li Q, Li S, Huo X, Li D (2020) Statistical analysis and countermeasures of major power customer loss. *IOP Publishing Ltd*, 2020, pp. 1-8. https://doi.org/10.1088/1755-1315/453/1/012057.

[5] Maw M, Haw S C, Ho C K (2021) Utilizing data sampling techniques on algorithmic fairness for customer churn prediction with data imbalance problems. *F1000Research*, 10, pp. 988. https://doi.org/10.12688/f1000research.72929.1.

[6] Tavassoli S, Koosha H (2022) Hybrid ensemble learning approaches to customer churn prediction. *Kybernetes*, 51 pp. 1062-1088. https://doi.org/10.1108/K-04-2020-0214.

[7] Zhu B, Qian C, Pan X (2020) A Trajectory-based deep sequential method for customer churn prediction. *ICMLT '20: Proceedings of the 2020 5th International Conference on Machine Learning Technologies*, 2020, pp. 114-118. https://doi.org/10.1145/3409073.3409083.

[8] Geiler L, Affeldt S, Nadif M (2022) An effective strategy for churn prediction and customer profiling. *Data & Knowledge Engineering*, 142, pp. 102100. https://doi.org/10.1016/j.datak.2022.102100.

[9] Karvana K G M, Yazid S, Syalim A, Mursanto P (2019) Customer churn analysis and prediction using data mining models in banking industry. *2019 International Workshop on Big Data and Information Security (IWBIS), Bali, Indonesia*, pp. 33-38. https://doi.org/10.1109/IWBIS.2019.8935884.

[10] Rahman M, Kumar V (2020) Machine learning based customer churn prediction in banking. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India*, pp. 1196-1201. https://doi.org/10.1109/ICECA49313.2020.9297529.

[11] Ishaq A, Sadiq S, Umer M, Ullah S, Mirjalili S, Rupapara V, Nappi M (2021) Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. *IEEE*

Access, 9, pp. 39707-39716. https://doi.org/10.1109/ACCESS.2021.3064084.

[12] Farooq S (2021) Comparison of data-driven landslide susceptibility assessment using weight of evidence, information value, frequency ratio and certainly factor methods. *Acta Geodynamica et Geomaterialia*, 18, pp. 301-317. https://doi.org/10.13168/AGG.2021.0021.

[13] Koo J, Paik S, Lee K (2021) Reverb conversion of mixed vocal tracks using an end-to-end convolutional deep neural network. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada*, pp. 81-85. https://doi.org/10.1109/ICASSP39728.2021.9414038.

[14] Lel T E, Ahsan M, Haider J (2023) Detecting COVID-19 from chest x-rays using convolutional neural network ensembles. *Computers*, 12, pp. 105. https://doi.org/10.3390/computers12050105.

[15] Bhatia Y, Bajpayee A, Raghuvanshi D, Mittal H (2019) Image captioning using Google's Inception-resnet-v2 and recurrent neural network. *2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India*, pp. 1-6. https://doi.org/10.1109/IC3.2019.8844921.

[16] Bhandari V, Londhe N D, Kshirsagar G B (2023) Compact temporal dilated convolution with Channel-wise attention and cost sensitive learning for Single trial P300 detection. *Biomedical Signal Processing and Control*, 85, pp. 1-13. https://doi.org/10.1016/j.bspc.2023.104924.

[17] https://www.kaggle.com/datasets

[18] Shadman Roodposhti M, Aryal J, Lucieer A, Bryan B A (2019) Uncertainty assessment of hyperspectral image classification: deep learning vs. random forest. *Entropy*, 21, pp. 78. https://doi.org/10.3390/e21010078.

[19] Zulfiqar H, Yuan S S, Huang Q L, Sun Z J, Dao F Y, Yu X L, Lin H (2021) Identification of cyclin protein using gradient boost decision tree algorithm. *Computational and Structural Biotechnology Journal*, 19, pp. 4123-4131. https://doi.org/10.1016/j.csbj.2021.07.013.

[20] Bhattacharya S, S S R K, Maddikunta P K R, Kaluri R, Singh S, Gadekallu T R, Alazab M, Tariq U (2020) A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. *Electronics*, 9, pp. 219-. https://doi.org/10.3390/electronics9020219.

[21] Zhang Q, Zhang C, Ni J, Wang X, Zhang Y (2021) Data sensitivity measurement and classification model of power IOT based on information entropy and BP neural network. *Journal of Physics: Conference Series*, 1848, pp. 1-6. https://doi.org/10.1088/1742-6596/1848/1/012107.