



ISSN: 0067-2904

Enhancing Customer Retention through Deep Learning and Imbalanced Data Techniques

Manal Loukili^{1*}, Fayçal Messaoudi², Mohammed El Ghazi³

¹National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco

²National School of Business and Management, Sidi Mohamed Ben Abdellah University, Fez, Morocco

³Superior School of Technology, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Received: 5/1/2023

Accepted: 5/5/2023

Published: 30/5/2024

Abstract

Accurately predicting customer churn is considered crucial by businesses in order to take proactive measures to retain their customers and avoid financial losses. In this paper, a customer churn prediction model is proposed that incorporates deep neural networks and imbalanced data techniques. The approach involves applying oversampling and undersampling methods to address class imbalances in the dataset. The model's performance is evaluated using various evaluation metrics and compared to other methods. The results demonstrate that superior performance in predicting customer churn is achieved by the proposed model compared to traditional statistical methods and deep neural networks without imbalanced data techniques. Important implications for businesses seeking to reduce customer churn and improve customer retention are provided by our findings. By using the suggested model, customers can be retained, and financial losses can be avoided proactively.

Keywords: Deep Neural Networks, Churn prediction, E-marketing, Oversampling, Undersampling.

1. Introduction

Customer churn, or customer turnover, is a common business problem that refers to the loss of customers over a given period of time [1]. It is a critical issue for businesses, as it can lead to financial losses and a decline in revenue. Customer churn can result from various factors, such as poor customer service, product quality issues, or the availability of better alternatives on the market.

In recent years, customer churn has become an increasingly significant concern for businesses due to the growing competition in many industries [2]. With the rise of e-commerce and online shopping, customers have more options than ever before, making it easier for them to switch to a competitor if they are not satisfied with a product or service.

To address this problem, businesses need to accurately predict which customers are most likely to churn and take proactive measures to retain them. Customer churn prediction is, therefore, an essential task for businesses seeking to improve customer retention and reduce financial losses.

*Email: manal.loukili@usmba.ac.ma

Traditional approaches to customer churn prediction have relied on statistical methods such as logistic regression or decision trees [3]. However, these methods have limitations, such as the need for manual feature engineering, which can be time-consuming and require domain expertise. Additionally, these methods may not be able to capture the complex relationships between customer behavior and churn.

Recently, deep learning approaches, such as deep neural networks (DNNs), have gained popularity in predictive modeling due to their capability to automatically learn and extract features from raw data [4]. DNNs have achieved state-of-the-art performance in numerous areas, such as natural language processing (NLP), computer vision, and predictive modeling [5]. In customer churn prediction, DNNs can help businesses identify patterns and relationships in large and complex datasets that may not be apparent using traditional statistical methods.

However, the effectiveness of DNNs in customer churn prediction has not been extensively explored, and there is a gap in the literature regarding their application to this problem. Additionally, imbalanced data, where the number of observations in one class significantly outweighs the other, is a common issue in customer churn prediction [6]. Imbalanced data can result in biased models that perform poorly in predicting the minority class. To address this issue, oversampling and undersampling algorithms can be applied to balance the class distribution [7].

Therefore, in this paper, we propose a customer attrition prediction model based on DNNs and imbalanced data techniques. The objective of this study is to demonstrate the efficiency of this approach in improving the accuracy of churn prediction. By exploring the potential of DNNs and imbalanced data techniques in customer churn prediction, we aim to provide new insights into the application of deep learning to real-world business problems.

The rest of this paper is structured as follows: In Section 2, a comprehensive review is provided of the relevant literature on the application of deep neural networks, imbalanced data techniques, and their use in predicting customer churn. The methodology adopted for this study is described in Section 3. In Section 4, the results and discussion of the model's performance, along with its implications and limitations, are presented. Finally, in Section 5, the paper is concluded by summarizing the main findings and their practical applications, as well as by suggesting directions for future research.

2. Literature Review

In the field of customer churn prediction, many researchers have proposed various machine learning and deep learning methods to identify which customers are likely to churn and to determine the key factors that cause them to do so.

Machine learning is a method of teaching computers to learn from data without being explicitly programmed. It is a subfield of artificial intelligence that involves the development of algorithms and statistical models that allow computers to improve their performance on a task by learning from examples or experience. These algorithms can be trained to make predictions, recognize patterns, classify data, and make decisions based on input data.

Deep neural networks (DNNs) are a form of machine learning technique inspired by the structure and function of the human brain [8]. They consist of multiple layers of interconnected artificial neurons, which process and transmit information through weighted

connections. DNNs have the potential to learn and extract features from raw data, making them effective across a broad range of applications.

In [9], the authors compared the performance of different machine learning models for predicting customer churn, which were balanced random forest, random forest, and logistic regression. They utilized principal component analysis (PCA) for feature selection. The results of their study showed that among these algorithms, the logistic regression model had the best performance, with an area under the curve (AUC) score of 0.86, which was higher than the other models.

In another study [10], the authors developed a machine learning model that is based on the customer's activity patterns. The model focuses on analyzing the customer's activity by looking for the average length of inactivity and the frequency of inactivity. The authors proposed that this method could also be applied to other areas for predicting customer churn.

The authors in [11] pointed out that the analysis of customer churn is essential for telecommunications companies to retain valuable clients. They emphasized that a more accurate customer churn prediction model is crucial for the decision-making process of customer retention. In their study, they employed a support vector machine (SVM) model because it is more precise and efficient for this task. The SVM model is able to solve samples in a low-dimensional space, which is linearly inseparable in a two-dimensional space. Nevertheless, the authors also acknowledged that one of the limitations of the suggested model is the complexity of measuring unsubscribed customers.

In [12], the author conducted a study in which he compared two prediction methods for customer churn, namely SMOTE and Deep Belief Network (DBN), with cost-sensitive learning approaches such as focal loss and weighted loss. The study results showed that focal loss and weighted loss performed better overall in comparison to SMOTE and DBN. These cost-sensitive learning techniques are especially useful when the cost of false negatives (misclassifying a positive sample) is much higher than the cost of false positives (misclassifying a negative sample), which is often the case with customer churn.

In [13], to predict customer churn, the authors applied deep learning models and 10-fold cross-validation methods to evaluate the prediction accuracy of the model. The results yielded an AUC score of 0.89, indicating a good performance. However, this score can be enhanced further by using additional features, more complex models, and running more iterations of training.

In the paper [14], the authors employed machine learning and deep learning approaches to identify customers at risk of unsubscribing and to determine the key factors that cause them to do so. They used multiple algorithms for prediction and comparison, and the results of the experiment showed that the Random Forest model performed the best among all the algorithms, followed by the Convolutional Neural Network (CNN) and Multi-Layer Perceptron (MLP) deep learning models. The study results highlighted the prediction models that are successful at identifying probable churners with the greatest accuracy, as well as the key factors that influence churn.

In the domain of predictive modeling, DNNs have achieved leading performance in various tasks such as image classification, speech recognition, and customer churn prediction. Several studies have demonstrated the effectiveness of DNNs in predicting customer churn,

with some achieving higher accuracy compared to traditional statistical methods [15], [16], [17], and [18].

Imbalanced data, where the number of observations in one class significantly outweighs the other, is a common issue in predictive modeling. This can lead to poor performance of classification algorithms, as they tend to prioritize the majority class. To address this issue, oversampling and undersampling techniques can be applied to balance the class distribution. Oversampling involves augmenting the number of minority-class observations to balance the class distribution. This is generally done by the generation of synthetic samples of the minority class or by randomly replicating existing minority class samples [19]. The goal of oversampling is to give the minority class more representation in the data so that the classifier can better learn its characteristics.

Undersampling involves reducing the number of majority-class observations to balance the class distribution [20]. This is achieved by choosing a random subset of the samples from the majority class or by using clustering algorithms to identify a representative subset of the majority class. The main purpose of undersampling is to reduce the influence of the majority class on the classifier so that it can better learn the characteristics of the minority class.

Several studies have investigated the use of imbalanced data sampling methods in customer churn prediction [21], [22], and [23]. Some have found that oversampling and undersampling can improve the performance of classification algorithms [24], [25], while others have reported mixed results.

3. Methodology

In order to improve customer churn prediction using deep neural networks and imbalanced data techniques, the steps outlined below (Figure 1) were followed in this study:

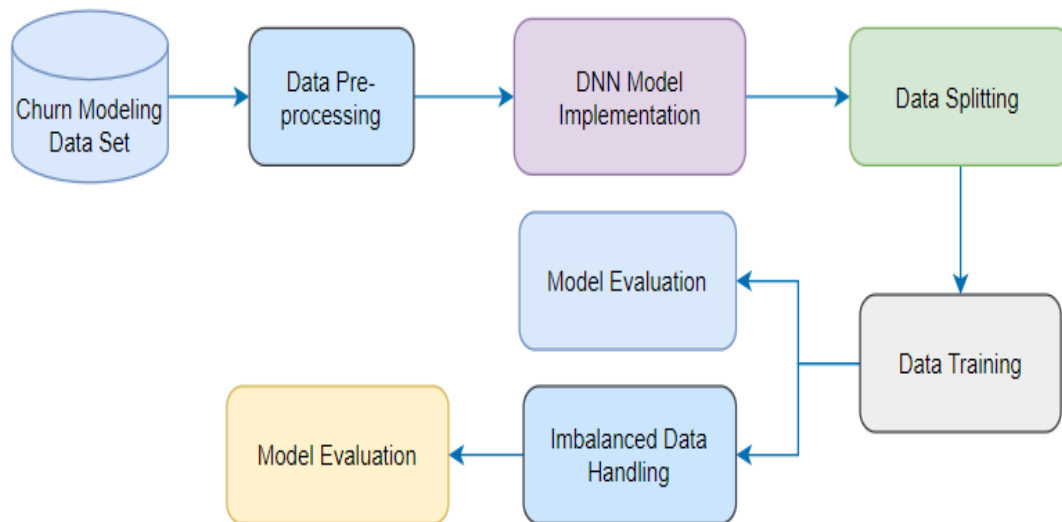


Figure 1: Steps of the adopted methodology

3.1 Data Pre-processing

The Churn Modeling data set was used, which consists of 10,000 rows and 11 columns containing details about a bank's customers (including: Age, CreditScore, Gender, Tenure, Geography, Balance, IsActiveMember, HasCrCard, NumOfProducts, and EstimatedSalary) and the target variable indicating if the customer has left the bank (closed his account) or remains a customer (Table 1).

Table 1: The data set used

	Cre ditS core	Geogr aphy	Gender	Age	Ten ure	Balance	NumO fProdu cts	HasCr Card	IsActiv eMem ber	Estima tedSal ary	Ex ited
0	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0
...
9995	771	France	Male	39	5	0.00	2	1	0	96270.64	0
9996	516	France	Male	35	10	57369.61	1	1	1	101699.77	0
9997	709	France	Female	36	7	0.00	1	0	1	42085.58	1
9998	772	Germa ny	Male	42	3	75075.31	2	1	0	92888.52	1
9999	792	France	Female	28	4	130142.79	1	1	0	38190.78	0

Next, the distributions of the different variables according to the target variable were visualized in order to understand the data set and determine the variables that influence the target variable, as shown in Figures 2 and 3.

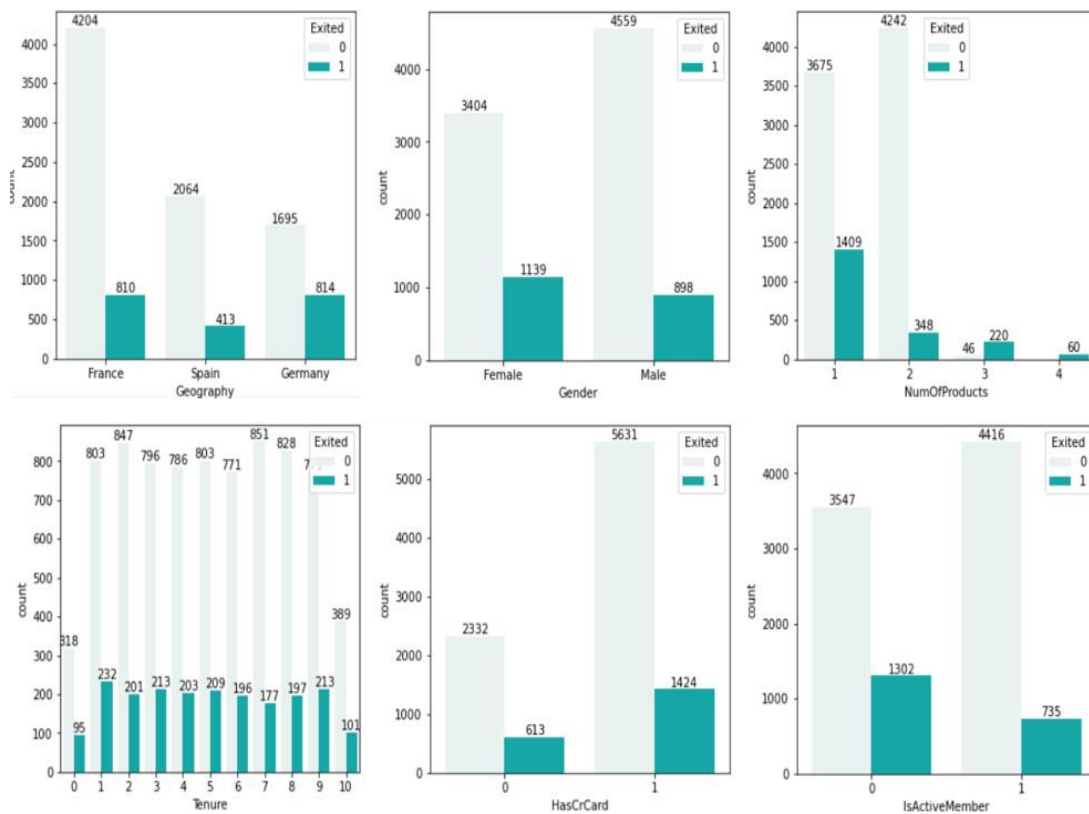


Figure 2: Bar plots of discrete variables against the target variable

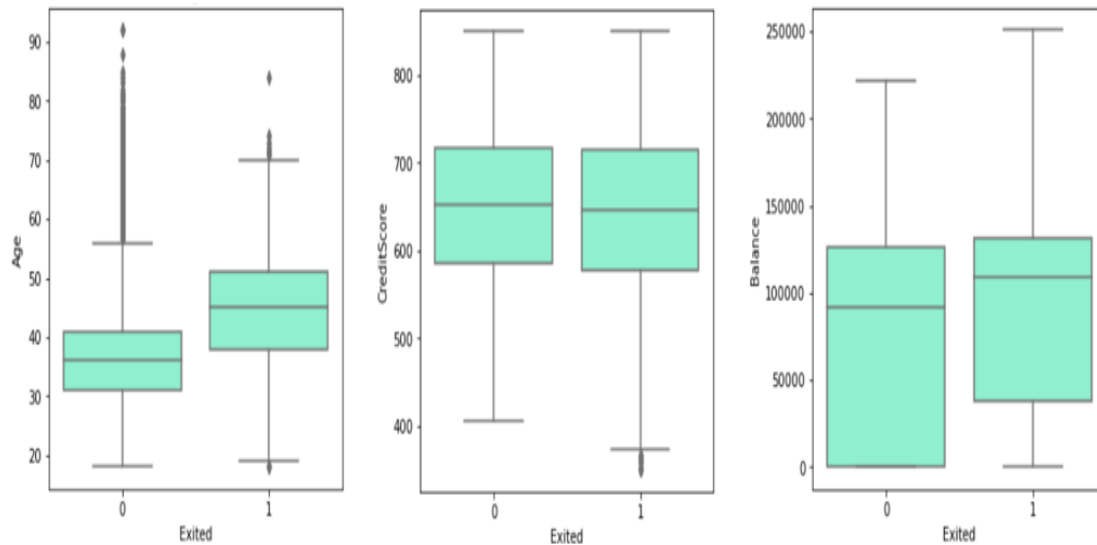


Figure 3: Box plots of continuous variables against the target variable

Several preprocessing steps were performed on the data before modeling. The features were standardized to have zero mean and unit variance; the target labels were encoded as integers between 0 and n-1 classes; and one-hot encoding was used to convert categorical features into a numeric array. The correlation between features was also examined to determine any redundant or strongly correlated variables, as shown in Figure 4.

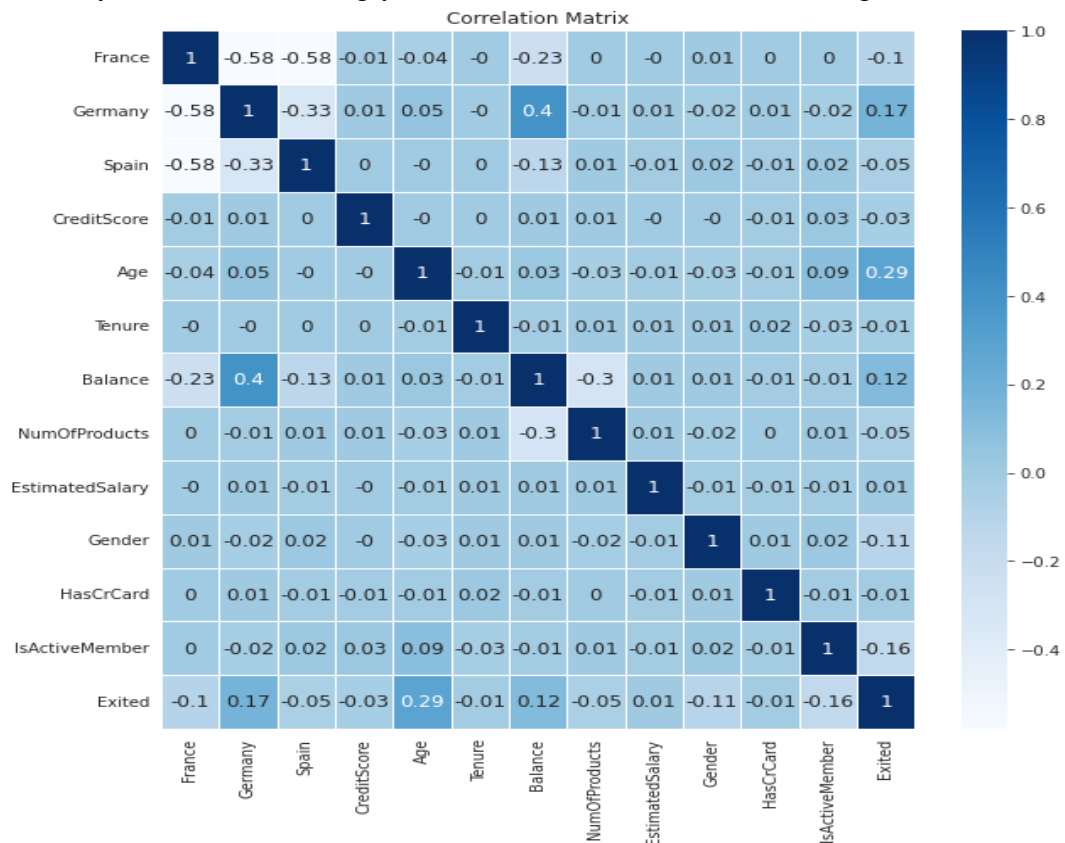


Figure 4: The correlation matrix

3.2 Predictive Model

For the predictive model, a DNN with a sequential model architecture was used. The DNN model consisted of 11 layers, including 8 fully connected dense layers, 3 dropout layers for regularization, and an output layer with a sigmoid activation function with 1 unit, as the target variable was binary (either 0 or 1). Figure 5 presents the architecture of the model.

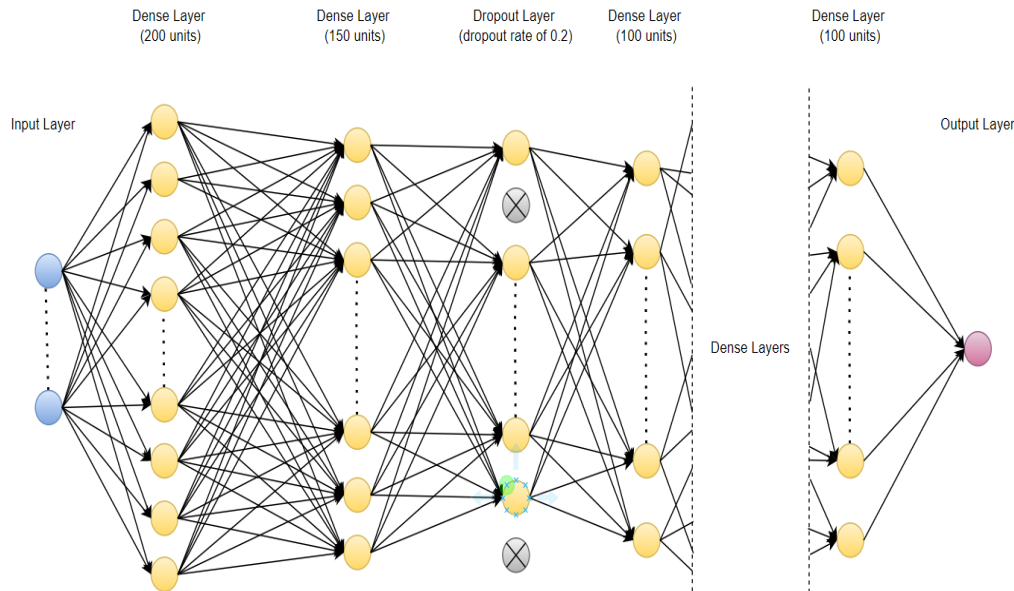


Figure 5: DNN model architecture

The layers of the model are the following:

- Input layer: 12-dimensional input vector.
- Dense layer 1: 200 neurons with ReLU activation function.
- Dense layer 2: 150 neurons with ReLU activation function.
- Dropout layer 1: applies dropout of 0.2.
- Dense layer 3: 100 neurons with ReLU activation function.
- Dense layer 4: 100 neurons with ReLU activation function.
- Dropout layer 2: applies dropout of 0.2.
- Dense layer 5: 100 neurons with ReLU activation function.
- Dense layer 6: 100 neurons with ReLU activation function.
- Dense layer 7: 100 neurons with ReLU activation function.
- Dropout layer 3: applies dropout of 0.2.
- Dense layer 8: 100 neurons with ReLU activation function.
- Output layer: 1 neuron with sigmoid activation function.

The model was compiled by applying the Adam optimizer (adaptive moment estimation) and binary cross-entropy loss function, with mean squared error and accuracy being used as metrics to evaluate the model's performance.

3.3 Training

The DNN for customer attrition prediction was trained using the fit function from the Keras library. The data set was divided into training and test sets by means of the `train_test_split` function from the `sklearn.model_selection` library, with a test set size of 10% and the random state set to 42. The model was then trained on the training set using a batch size of 10 and 50 epochs. Validation data (the test set) was also included in the training process to assess the model's generalization performance.

3.4 Evaluation

The performance of the DNN in predicting customer churn was evaluated using the evaluate function from the Keras library on the test set. This function provided the model's mean squared error (MSE), accuracy, and loss metrics.

- The MSE metric measures the average squared difference between the predicted and actual values for the target variable (a lower MSE indicates a better fit of the model to the data). MSE is written as follows (1):

$$MSE = (1/n) * \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{1}$$

Where:

n = total number of observations

y_i = actual value of the target variable for i^{th} observation

\hat{y}_i = predicted value of the target variable for i^{th} observation

- The accuracy metric measures the percentage of correct predictions made by the model (a higher accuracy indicates a better fit of the model to the data). Accuracy is written in the following form (2):

$$Accuracy = (number\ of\ correct\ predictions / total\ number\ of\ predictions) * 100 \tag{2}$$

- The loss metric measures the difference between the predicted and actual values for the target variable (a lower loss indicates a better fit of the model to the data).

$$Loss = - \sum_{i=1}^n (y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i)) \tag{3}$$

Where:

n = total number of observations

y_i = actual value of the target variable for i^{th} observation (0 or 1)

\hat{y}_i = predicted probability of the target variable being 1 for i^{th} observation

The loss is calculated based on the cross-entropy between the actual and predicted distributions of the target variable.

Applying imbalanced data techniques to the model would involve first balancing the class distribution of the data set using the desired technique. The resulting balanced data set would then be used to train and evaluate the model using the same steps as described above.

3.5 Imbalanced Data Handling

Due to the class imbalance in the data, as shown in Figure 6, several oversampling and undersampling techniques were applied to the training data in order to address the issue of class imbalance in the data set.

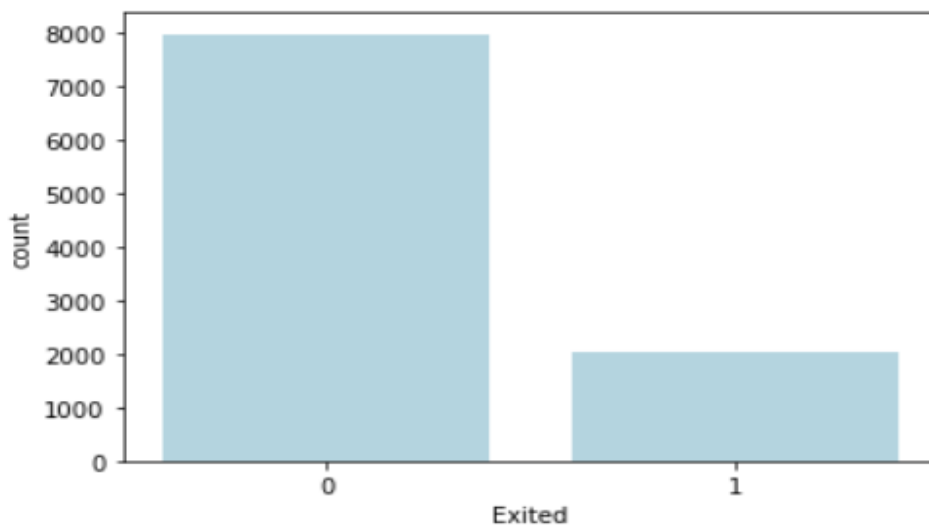


Figure 6: Unbalanced target variable classes

The oversampling techniques used are Synthetic Minority Oversampling Technique, RandomOverSampler, BorderlineSMOTE, Adaptive Synthetic, KMeansSMOTE, and Support Vector Machine SMOTE.

- Synthetic Minority Oversampling Technique (SMOTE): This is a popular oversampling technique that generates synthetic samples of the minority class by interpolating between existing minority class samples [26]. The synthetic samples are generated by randomly choosing two minority-class samples and creating a new sample at a random point along the line connecting the two samples.
 - RandomOverSampler: This is a simple oversampling technique that randomly selects samples from the minority class and replicates them until the class distribution is balanced [27].
 - BorderlineSMOTE: This is an oversampling technique that generates synthetic samples of the minority class based on the border between the minority and majority class samples. It creates synthetic samples by randomly selecting a minority class sample and a majority class sample that is closest to the minority sample and generating a new sample at a random point along the line connecting the two samples [28].
 - Adaptive Synthetic (ADASYN): This is an oversampling technique that generates synthetic samples of the minority class based on the density of the samples in the feature space. It creates synthetic samples by selecting a minority class sample and finding the k-nearest neighbors of that sample in the feature space [29]. A new sample is then generated at a random point along the line connecting the selected sample and one of its neighbors, with the probability of choosing a neighbor based on the density of the samples in the neighborhood.
 - KMeansSMOTE: This is an oversampling technique that uses the K-means clustering algorithm to cluster the minority class samples and generate synthetic samples within the clusters [30]. It first clusters the minority class samples into k clusters using K-means and then generates synthetic samples within each cluster by randomly selecting two samples and creating a new sample at a random point along the line connecting the two samples.
 - Support Vector Machine SMOTE (SVMSMOTE): This is an oversampling technique that uses the support vector machine (SVM) algorithm to generate synthetic samples of the minority class [31]. It first trains an SVM classifier on the minority class samples and then generates synthetic samples by sampling from the decision boundary defined by the SVM.
- The undersampling techniques applied are ClusterCentroids, AllKNN, NeighbourhoodCleaningRule, and RandomUnderSampler.
- ClusterCentroids: This technique involves clustering the majority of class observations into k clusters using an unsupervised learning algorithm such as k-means [32]. The centroids of these clusters are then selected as representative samples of the majority class, and the rest of the majority class observations are discarded. This results in a balanced data set with a reduced number of majority-class observations.
 - AllKNN: This technique involves selecting a subset of the majority class observations that are nearest to the minority class observations based on a distance metric such as Euclidean distance. The majority class observations that are farther away from the minority class observations are discarded [33].
 - NeighbourhoodCleaningRule: This technique involves selecting a subset of the majority-class observations that are nearest to the minority-class observations based on a distance metric such as Euclidean distance. The majority class observations that are farther away from the minority class observations are discarded [34]. In addition, the majority class observations that are nearest to the minority class observations are removed if they are not

similar to the minority class observations, based on a similarity measure such as the Jaccard coefficient.

- **RandomUnderSampler:** This technique involves randomly selecting a subset of the majority class observations to balance the class distribution. The size of the subset is determined by the desired balance between the classes [35].

3.6 Model Comparison

The effectiveness of the DNN model combined with data sampling methods was evaluated by comparing the performance metrics of the DNN model before and after applying oversampling and undersampling techniques. Additionally, the performance of the DNN model was compared with that of a commonly used machine learning model, logistic regression, which is a simple linear model that can serve as a benchmark for comparison with more complex models like DNN. Good performance in predicting customer churn has been demonstrated in many cases [36, 37] by logistic regression. By comparing the performance of the proposed DNN-based model with a well-established benchmark, a more robust assessment of its effectiveness and practical utility was provided.

Furthermore, the different data sampling techniques used in conjunction with the DNN model were compared to select the best technique that yields the highest accuracy. The accuracy, precision, and recall of the DNN model were evaluated using several data sampling techniques. The resulting balanced dataset was then used to train and evaluate the model. The comparison of the DNN model with different data sampling techniques allows the selection of the most effective technique for improving the model's accuracy.

4. Results and Discussion

4.1 Results

In this paper, a deep neural network (DNN) was implemented and evaluated on the Churn Modeling data set using various imbalanced data techniques to address the class imbalance in the data. The results of the predictions are summarized in Tables 2 and 3.

Table 2: DNN model evaluation without data sampling techniques

Model	Accuracy	MSE	Loss
DNN	0.88	0.11	0.37

Table 3: Model evaluation with the different data sampling techniques

Over/Under -sampling method	Accuracy	MSE	Loss
SMOTE	0.88	0.08	0.31
RandomOverSampler	0.90	0.07	0.30
BorderlineSMOTE	0.88	0.08	0.38
ADASYN	0.89	0.08	0.30
KMeansSMOTE	0.88	0.08	0.34
SVMSMOTE	0.89	0.08	0.34
ClusterCentroids	0.91	0.07	0.28
AIKNN	0.90	0.07	0.31
NeighbourhoodCleaningRule	0.94	0.04	0.35
RandomUnderSampler	0.95	0.04	0.18

It was found that the DNN combined with the RandomUnderSampler technique outperformed the other sampling techniques with an accuracy of 0.95, a mean squared error (MSE) of 0.04, and a loss of 0.18. The DNN alone achieved an accuracy of 0.88, an MSE of 0.11, and a loss of 0.37. The oversampling techniques (SMOTE, RandomOverSampler, BorderlineSMOTE, ADASYN, KMeansSMOTE, and SVMSMOTE) produced mixed results, with some achieving higher accuracy and lower MSE and loss compared to the DNN alone, while others had lower accuracy and higher MSE and loss.

Furthermore, the performance of the DNN with and without imbalanced data techniques was compared to a baseline model using logistic regression. The DNN with RandomUnderSampler significantly outperformed the logistic regression model, with an accuracy of 0.95 compared to 0.78 for the logistic regression model. These results indicate that the use of DNNs and imbalanced data techniques can improve the accuracy of customer churn prediction compared to traditional machine learning techniques such as logistic regression.

4.2 Discussion

In sum, the DNN combined with imbalanced data techniques was found to effectively improve the accuracy of customer churn prediction. The use of the RandomUnderSampler technique, in particular, appears to be effective in improving the performance of the DNN on this data set.

The results of this study have important implications for businesses that rely on customer retention. The findings demonstrate that the use of DNNs and imbalanced data techniques can significantly improve the accuracy of customer churn prediction, which can inform the development of more accurate and effective customer churn prediction models.

One limitation of this study is that only a limited number of oversampling and undersampling techniques were tested, and further research is needed to determine the effectiveness of these techniques on other data sets. Another limitation is that the study did not investigate the interpretability of the DNN models, which is an important consideration for practical applications of these models.

5. Conclusion

In conclusion, this study has demonstrated the effectiveness of using deep neural networks (DNNs) and imbalanced data techniques for predicting customer turnover in the banking industry. The use of the Churn Modeling data set and various oversampling and undersampling techniques successfully addressed the class imbalance in the data, resulting in improved accuracy of customer churn prediction.

The DNN model combined with the RandomUnderSampler technique achieved the highest accuracy among all the imbalanced data techniques, with an accuracy of 0.95, a mean squared error (MSE) of 0.04, and a loss of 0.18. This suggests that the RandomUnderSampler technique may be particularly effective in addressing the class imbalance in this data set and improving the performance of the DNN.

Future research directions could include exploring additional imbalanced data techniques, such as hybrid techniques that combine oversampling and undersampling, as well as data synthesis, data augmentation, and cost-sensitive learning. Additionally, investigating different

neural network architectures, such as convolutional neural networks or recurrent neural networks, could lead to further improvements in accuracy.

This study has made a valuable contribution to the field of customer churn prediction in the banking industry. By demonstrating the effectiveness of DNNs and imbalanced data techniques, businesses can improve their customer retention strategies and ultimately improve their bottom line.

References

- [1] M. Loukili, F. Messaoudi and M. El Ghazi, "Supervised Learning Algorithms for Predicting Customer Churn with Hyperparameter Optimization," *Int. J. Advance Soft Compu. Appl*, vol. 14, no. 3, pp. 49-63, 2022. doi: 10.15849/IJASCA.221128.04
- [2] Y. Richter, E. Yom-Tov and N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups," in *Proceedings of the 2010 SIAM international conference on data mining*, pp. 732-741, Columbus, Ohio, USA, 2010. Doi: 10.1137/1.9781611972801.64
- [3] A. De Caigny, K. Coussement and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *European Journal of Operational Research*, vol. 269, no. 2, pp. 760-772, 2018.
- [4] H. Liang, X. Sun, Y. Sun and Y. Gao, "Text feature extraction based on deep learning: a review," *EURASIP journal on wireless communications and networking*, vol. 2017, no. 1, pp. 1-12, 2017. doi: 10.1186/s13638-017-0993-1
- [5] W. He, "Load forecasting via deep neural networks," *Procedia Computer Science*, vol. 122, pp. 308-314, 2017. doi: 10.1016/j.procs.2017.11.374
- [6] Y. Li, J. Wei, K. Kang and Z. Wu, "An efficient noise-filtered ensemble model for customer churn analysis in aviation industry," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 2, pp. 2575-2585, 2019.
- [7] R. Mohammed, J. Rawashdeh and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *2020 11th international conference on information and communication systems (ICICS)*, pp. 243-248, IEEE, 2020.
- [8] J. Moolayil, *An introduction to deep learning and keras, Learn Keras for Deep Neural Networks*, Apress, Berkeley, CA, 2019, pp. 1-16. doi: 10.1007/978-1-4842-4240-7
- [9] N. Kumar and C. Naik, "Comparative Analysis of Machine Learning Algorithms for Their Effectiveness in Churn Prediction in the Telecom Industry," *International Research Journal of Engineering and Technology*, vol. 4, pp. 485-489, 2017.
- [10] A. Bharat, "Consumer Engagement Pattern Analysis Leading to Improved Churn Analytics: An Approach for Telecom Industry," *Advances in Intelligent Systems and Computing*, pp. 203-211, 2018. doi: 10.1007/978-981-13-1274-8_16
- [11] X. Zhang, Z. Zhang, D. Liang D and H. Jin, "A Novel Decision Tree Based on Profit Variance Maximization Criterion for Customer Churn Problem," in *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, pp. 20-23, 2018. doi: 10.1109/IHMSC.2018.00013.
- [12] N. Nam and 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, pp. 29-35, 2021. doi: 10.12720/jait.12.1.29-35
- [13] R. Andrews, R. Zacharias, S. Antony and M. M. James, "Churn Prediction in Telecom Sector Using Machine Learning," *International Journal of Information Systems and Computer Sciences*, vol. 8, pp. 132-134, 2019. doi: 10.30534/ijiscs/2019/31822019
- [14] A. Siddika, A. Faruque and A. K. M. Masum, "Comparative Analysis of Churn Predictive Models and Factor Identification in Telecom Industry," in *2021 24th International Conference on Computer and Information Technology ICCIT*, pp. 1-6, 2021. doi: 10.1109/ICCIT5 4785.2021.9689881
- [15] A. De Caigny, K. Coussement, K. W. De Bock and S. Lessmann, "Incorporating textual information in customer churn prediction models based on a convolutional neural network," *International Journal of Forecasting*, vol. 36, no. 4, pp. 1563-1578, 2020. doi: 10.1016/j.

- ijforecast.2019.03.029
- [16] O. Çelik, and U. O. Osmanoglu, "Comparing to techniques used in customer churn analysis," *Journal of Multidisciplinary Developments*, vol. 4, no. 1, pp. 30-38, 2019.
- [17] G. G. Mena, A. De Caigny, K. Coussement, K. W. De Bock and S. Lessmann, "Churn prediction with sequential data and deep neural networks, a comparative analysis," *arXiv preprint arXiv:1909.11114*, 2019.
- [18] M. Saghir, Z. Bibi, S. Bashir and F. H. Khan, "Churn prediction using neural network based individual and ensemble models," in *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 634-639, IEEE, 2019.
- [19] V. A. Fajardo, D. Findlay, C. Jaiswal, X. Yin, R. Houmanfar, H. Xie and D. B. Emerson, "On oversampling imbalanced data with deep conditional generative models," *Expert Systems with Applications*, 169, ref. 114463, 2021. doi: 10.1016/j.eswa.2020.114463
- [20] J. Hoyos-Osorio, A. Alvarez-Meza, G. Daza-Santacoloma, A. Orozco-Gutierrez and G. Castellanos-Dominguez, "Relevant information undersampling to support imbalanced data classification," *Neurocomputing*, vol. 436, pp. 136-146, 2021.
- [21] 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.
- [22] K. S. Karuppaiah and N. G. Palanisamy, "Heterogeneous ensemble stacking with minority upliftment (HESMU) for churn prediction on imbalanced telecom data," *Materials Today: Proceedings*, 2021. doi: 10.1016/j.matpr.2020.12.893
- [23] A. Bhattarai, E. Shrestha, and R. P. Sapkota, "Customer churn prediction for imbalanced class distribution of data in business sector," *Journal of Advanced College of Engineering and Management*, vol. 5, pp. 101-110, 2019.
- [24] Z. P. Agusta and A. Adiwijaya, "Modified balanced random forest for improving imbalanced data prediction," *International Journal of Advances in Intelligent Informatics*, vol. 5, no. 1, pp. 58-65, 2019.
- [25] S. Shumaly, P. Neysaryan and Y. Guo, "Handling class imbalance in customer churn prediction in telecom sector using sampling techniques, bagging and boosting trees," in *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 82-87, IEEE, 2020.
- [26] D. Elreedy and A. F. Atiya, "A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance," *Information Sciences*, vol. 505, no. C, pp. 32-64, 2019. doi: 10.1016/j.ins.2019.07.070
- [27] F. de la Bourdonnaye and F. Daniel, "Evaluating resampling methods on a real-life highly imbalanced online credit card payments dataset," *arXiv preprint, arXiv:2206.13152*, 2022.
- [28] S. Smiti and M. Soui, M. "Bankruptcy prediction using deep learning approach based on borderline SMOTE," *Information Systems Frontiers*, vol. 22, no. 5, pp. 1067-1083, 2020.
- [29] J. Liu, Y. Gao and F. Hu, "A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM," *Computers & Security*, vol. 106, no. C, 2021. doi: 10.1016/j.cose.2021.102289
- [30] J. Fonseca, G. Douzas and F. Bacao, "Improving imbalanced land cover classification with K-Means SMOTE: Detecting and oversampling distinctive minority spectral signatures," *Information*, vol. 12, no. 7, 266. 2021. doi: 10.3390/info12070266
- [31] R. Malhotra, V. Agrawal, V. Pal and T. Agarwal, "Support Vector based Oversampling Technique for Handling Class Imbalance in Software Defect Prediction," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 1078-1083, IEEE, 2021.
- [32] B. R. Silva, R. J. Silveira, M. G. da Silva Neto, P. C. Cortez and D. G. Gomes, "A comparative analysis of undersampling techniques for network intrusion detection systems design," *Journal of Communication and Information Systems*, vol. 36, no. 1, pp. 31-43, 2021.
- [33] M. Mahbobi, S. Kimiagari and M. Vasudevan, "Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks," *Annals of Operations Research*, pp. 1-29, 2021. doi: 10.1007/s10479-021-04114-z
- [34] D. Javale, P. Pillai, P. Patel and S. Jagtap, "A Comparison of Oversampling and Undersampling

- Methods for Predicting Air Quality in Metropolitan Region," in *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pp. 1615-1620, IEEE, 2022.
- [35] M. Hanafy and R. Ming, "Improving imbalanced data classification in auto insurance by the data level approaches," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021.
- [36] E. Stripling, S. vanden Broucke, K. Antonio, B. Baesens and M. Snoeck, "Profit maximizing logistic regression modeling for customer churn prediction," in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1-10, IEEE, 2015.
- [37] H. Jain, A. Khunteta and S. Srivastava, "Churn prediction in telecommunication using logistic regression and logit boost," *Procedia Computer Science*, vol. 167, pp. 101-112, 2020.