An Analysis of Customer Churn Prediction in Different Business Industries

Zhengyang Zhao[a]
Electronic Information Engineering, South China Agricultural University, Guangzhou City, China

Keywords: Artificial Intelligence, Machine Learning, Deep Learning.

Abstract: In this article, the currently deployed forecasting techniques are reviewed. Churn is widely used for areas such as web services, gaming and insurance. However, since it is vastly used to improve predictability in various industries, there is a great deal of variation in its definition and usage. This paper categorises the traditional methods of machine learning and deep learning, presents a number of papers related to these two technologies, and discusses and analyses the papers in order to provide more academics with a clear understanding of how these two technologies are used in different industries. The paper brings together definitions of froth in the following areas as business management, Information and communication technology (ICT) and newspaper industry, and explains the differences between them. On the basis of this, churn loss, attribute engineering and predictive modelling are categorised and explained. This study can be conducted by debris integration studies in industrial domains and selecting churn definitions and relevant models suitable for most interest to researchers.

1 INTRODUCTION

The term "customer churn" is commonly used to describe a customer's tendency to stop working with an organisation for a specific period of time or contract (Chandar, 2006). Preventing customer churn is critical when operating a service. In the past, the efficiency of customer acquisition related to the amount of repeat customers was favourable. However, with the globalisation of services and intense competition leading to market saturation, customer acquisition costs are rising rapidly (Verbraken, 2014).

For technology companies, customer persona profiling is a major challenge in the contemporary business environment (Ebiaredoh-Mienye, 2021). These companies always suffer heavy losses due to customer churn. Early identification of customer personality traits is important to minimise customer churn and develop loyal customers especially in case of misinformation (Awan, 2022). Many studies have been conducted in the past to analyse customer churn and develop strategies to reduce it. Online shopping platforms in particular have the advantage of being easily accessible through PC web pages or mobile apps, but conversely, this advantage can also be a disadvantage in terms of being easily seen and quickly left (Seo, 2023). Therefore, even a slight decrease in customer churn can lead to higher conversion rates, which can result in huge profits (Ahmed, 2024). For these reasons, predicting customer churn can be used as a way to increase the value of the company.

Customer Relationship Management (CRM) initially emerged as a business management approach to improve efficiency in areas such as marketing, sales and business administration, as well as to enhance organisational efficiency and customer value functions (Parvatiyar, 2001). It has been used to develop marketing strategies using personal and behavioural data of customers, particularly to meet individual and unique consumer needs (Shaw, 2001). Since then, a number of companies, taking full advantage of Information technology (IT), have begun to apply specialised techniques for customer acquisition, retention and selection (Kumar, 1996). With the integration of IT and CRM technologies, an increasing number of organisations are adopting these technologies in areas as diverse as data warehousing, online platforms and finance (Bose, 2002). Due to developments in big data, many data mining and machine learning solutions available to analyse this

[a] https://orcid.org/0009-0006-7441-2384

Zhao, Z.
An Analysis of Customer Churn Prediction in Different Business Industries.
DOI: 10.5220/0012972800004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 781-785
ISBN: 978-989-758-713-9

data, they can analyse the data and discover the underlying causes of customer churn. Moreover, they can be used to design customer retention strategies to minimise customer churn (Ullah, 2019). Nowadays, churn analysis has become an important strategy for personalised customer management, and studies have shown that improving retention of specific groups of existing customers is more beneficial than attracting new customers (Jahromi, 2010). Many studies have applied several deep learning model-based froth analysis techniques to services in the field of computer science (Lee, 2019; Zhang, 2017).

The rest of the paper is organised as follows: In section 2 introduces some methods done by other researchers and discusses how they use machine learning and deep learning to address customer churn in different industries. In Section 3, the paper discusses the results of researchers in various industries who have used these methods to detect customer churn in recent years. In Section 4, the paper presents its conclusions.

2 METHOD

2.1 Traditional Machine Learning-Based Algorithms

This paper explores the application of machine learning and deep learning to customer churn in non-contractual settings over the past decade. Machine learning offers robust capabilities for capturing nonlinear relationships among features, allowing it to discern varied effects based on different characteristics (Qiu, 2020). Deep learning, a sophisticated extension of machine learning and neural networks, has become increasingly popular for predicting customer churn. It differs significantly from traditional models and is often considered a distinct category. Deep learning models typically involve training on condensed sparse customer data or using fully connected neural networks. These networks often incorporate latent vectors extracted from autoencoders, linking these vectors with static data to predict churn effectively (Ahn, 2020).

2.1.1 Random Forest

Random Forest (RF) is an ensemble learning technique that enhances model performance by integrating multiple decision trees. This method clusters data into smaller groups, with each subset being used to train an individual decision tree (Nath, 2003). Each decision tree in RF is trained using randomly selected data points with replacement using a technique called bootstrap clustering. In addition, a random subset of the quality for each fork in the decision tree is chosen to be considered instead of all features. This ramps up generality of model and reduces overfitting.

To combine customer churn prediction and segmentation, Olah et al. proposed a churn prediction and customer segmentation framework (Olah, 2019). They used RF to predict customer churn and gain insight into the pivotal factors that contribute to customer churn, they identified the factors using an attribute selection classifier. Next, they extracted all the customer churn data that was properly predicted by the RF and performed a customer analysis to understand the similarities between these churned customers. Ultimately, based on the results of the analysis, some retention strategies and recommendations were made.

2.1.2 Decision Tree

Decision Tree (DT) is a tree structure similar to flowchart, where each internal node exhibits a test for an attribute, each branch exhibits the result of test, and leaf nodes represent final result or classification. They can be used for both classification and regression, and are created through a process called iterative partitioning, where data is repeatedly divided into subsets based on certain attribute values. The goal is to create a tree that accurately predicts the target variable, with the most important variables at the top of the tree. These algorithms differ in the way they choose attributes to partition the data and how they handle missing values and continuous variables. DT algorithms have many advantages: They easily visualise and understand, can handle numerical data, use a non-parametric approach and do not require a prior assumptions (Hassouna, 2016). Umayaparvathy et al. conducted a comparison for predicting confusion between Artificial Neural Network (ANN) and DT, and found that the decision tree-based method was more accurate than the neural network-based method (Umayaparvathy, 2012).

In a study by (Dahiya, 2015), decision trees were employed to predict customer churn, demonstrating superior performance over logistic regression. The decision trees achieved an impressive accuracy of 99.67% on a large dataset (Bamina, 2019). Another study found that XGBoost led in terms of accuracy, reaching 79.8%, and also scored the highest Area Under the Curve (AUC) at 58.2% for predicting customer churn.

In the study conducted by (Ullah, 2019), several machine learning techniques were used to classify customer data using annotated datasets. The aim was to evaluate which algorithm best categorises customers into frequent and non-frequent customer categories. The DT algorithm was used for classification. It was classified as an enthusiastic learning algorithm where the training data is generalised to classify new samples. This algorithm is an improved version of the original ID3 and C4.5, and is widely used in the literature to analyse data.

2.1.3 Support Vector Machine

Support Vector Machine (SVM) is a powerful machine learning model for classification and regression tasks. On the basis of the concept of a decision plane that defines boundaries of a decision. It works by mapping the input data into a high-dimensional feature space where hyperplanes can be used to discriminate between different classes. The hyperplane is chosen in such a way that the margin between two classes is maximised. This machine has many advantages, including the ability to handle high-dimensional data, efficiency in dealing with small datasets, and robustness to outliers.

More specifically, in the current study, they developed and implemented a hierarchical joint Bayesian model to predict intervals between events and the number of customer events using YouView data (Moral, 2022). When they attempted to classify customer status ("YES" vs. "NO" subscription customers), the results got using hierarchical joint Bayesian model parameter estimation outperformed the results obtained using the raw data from all machine learning methods, and in terms of accuracy, the SVM approach was the best performing overall, with 92% accuracy, 100% correct positive rate, and 14% false positive rate.

2.2 Deep Learning

Deep Learning (DL) is a more recent analytical approach to predicting disruption. According to Ian et al., it is part of machine learning (Goodfellow, 2016). Due to its increasing industries importance in recent times, it has become a separate academic field. It is also true for building models for flop prediction analyses. In their 2019 study, Lee, Eunju et al. demonstrated that DL models were more effective in predicting game flops compared to traditional methods (Lee, 2019). They enhanced prediction accuracy by integrating deep learning with traditional machine learning techniques, utilizing feature modification strategies such as memory and generalization. Zhang, Rong et al. also explored the effectiveness of deep learning versus traditional machine learning in predicting customer churn in the insurance sector (Zhang, 2017). They processed features specifically for deep learning applications and merged these insights with traditional models. Their findings showed that the deep learning-based churn prediction method outperformed conventional machine learning algorithms in terms of accuracy.

2.2.1 Artificial Neural Networks

ANN is an artificial intelligence system inspired by the human brain. It is composed of interconnected units known as nodes or neurons, which collaborate to process information. Each neuron receives input signals from other neurons and produces output signals to pass on to other neurons in the network. The input layer receives the data to be processed while the output layer produces the final result. The intermediate hidden layer performs various calculations and transformations on the data. It as been highly successful in solving complex problems that are difficult for traditional algorithms to handle. However, it requires large amounts of data and computational resources for training, and their performance may depend heavily on the quality of the data used for training.

The research by Arokia Panimalar and Krishnakumar is centered on creating a robust customer churn prediction model known as DFE-WUNB, designed to operate within cloud computing frameworks. This model leverages ANN for deep feature extraction, effectively addressing the intricate non-linear patterns found in telecommunications customer churn datasets. The DFE-WUNB model demonstrates superior accuracy in predicting customer churn compared to conventional methods (Panimalar, 2023).

2.2.2 Convolutional Neural Network

Convolutional Neural Network (CNN) is a deep learning algorithm primarily used for image recognition and classification tasks. Inspired by the structure and function of the human visual system, it is highly effective in recognising objects within the visual area. Firstly, there are convolutional layers: These layers apply a series of filters to the input image to create feature maps that capture different aspects of the image, such as edges, corners, or specific patterns. Secondly, activation layers: After each convolutional layer, an activation layer is usually applied to introduce nonlinearity to the network, allowing it to learn more complex patterns; Rectified Linear Unit (ReLU) is a common activation function used in WSNs. Third, clustering layers: These layers reduce the spatial size of the feature map, helping to reduce computational complexity, prevent overfitting, and make feature detection independent of size and orientation: After several layers of convolutional and clustering layers, the final set of layers in a CNN is usually one or more fully connected layers that perform high-level inference and categorise input image into predefined classes: The last layer of the network produces the output. The output can be a probability distribution of classes in a classification task or a bounding box in an object detection task.

CNNs have revolutionised various fields like computer vision and medical imaging. They are particularly effective because they are able to automatically learn and extract relevant features from raw data without the need to manually engineer features. In short, Improved RoCE Network (IRNs) are powerful deep learning models that excel at processing network-like data such as images, which are the cornerstone of modern image analysis and recognition systems. In a study by Ahmed et al. (Ahmed, 2019), Ahmed et al. utilized a DP approach and proposed a method called "TL-DeepE", which starts with TL (transfer learning) by tuning several pre-trained deep CNNs. They converted the TL dataset into a 2D image format. They then used these CNNs as base classifiers and Genetic Programming (GP) and AdaBoost as meta-classifiers. The accuracy of their method on the Orange and Cell2Cell datasets was 75.4% and 68.2% with an overall utilisation rate of 83% and 74%, respectively.

3 DISCUSSIONS

It can be confirmed that the modelling techniques favoured by different business domains are different. Companies in the gaming, social media and telecoms industries, which rely heavily on log data and have easy access to customer information, use deep learning techniques, which have relatively more applications for big data, and this is a fast-growing trend. For example, in the area of image recognition for social media platforms, many social media companies want to automatically tag user-uploaded images with relevant keywords to improve searchability and provide better content recommendations. They decided to use deep learning, specifically CNN, to recognise and categorise the content of these images. CNN is trained on large datasets of tagged images and can learn complex patterns and features directly from the data. When a new image is uploaded, the model is able to predict what is depicted in the image with high accuracy, even if the image is slightly different from the one in the training set. This deep learning approach enables social media platforms to optimise image tagging systems, leading to improved user experience and engagement.

For the financial and insurance industries, they use traditional machine learning models or analyses due to relatively small volume of log data and the small degree of variation in the information obtained from customers. For example, in the area of credit risk assessment, some financial firms (e.g., banks or credit unions) want to optimise their processes for assessing the creditworthiness of potential borrowers. Accurately predicting whether a borrower is likely to default on a loan is critical for financial institutions to manage risk and maintain profitability. Companies have historical data on loans, including various attributes of borrowers (e.g., income, employment status, credit history) and whether they will ultimately default on the loan. By using a VPPM to assess credit risk, financial firms can reduce the number of non-performing loans and improve the health of their overall loan portfolio.

In new tech companies have to deal with large amounts of complex data, so they will opt for high-cost deep learning models to facilitate data processing, giving them more optimistic expected returns. Whereas in traditional financial firms, they are more likely to choose lower-cost models to minimise costs as there is less variation in customer information. In addition, the developed algorithms should rely on more advanced hardware or transmission mechanisms to achieve higher processing speeds and more accurate identification capabilities (Deng, 2023; Sugaya, 2019).

4 CONCLUSIONS

This paper compares techniques for predictive analysis of user momentum using log data. In recent years, methods that use deep learning algorithms to analyse the prediction of user momentum have emerged. Deep learning algorithms outperform other algorithms. Unlike other modelling techniques, they are able to learn customer behavioural patterns from massive amounts of data through layers of stacked neuron structures. Therefore, applying this data to deep learning algorithms to generate latent features given the timestamp and large number of observations is expected to perform better than traditional fuzzy prediction models. Therefore, the reader needs to understand the shape of the datasets and apply the appropriate algorithms to solve the prediction problem.

REFERENCES

Awan, M. J., Khan, M. A., Ansari, Z. K., Yasin, A., & Shehzad, H. M. F. 2022. Fake profile recognition using big data analytics in social media platforms. *International Journal of Computer Applications in Technology*, *68*(3), 215.

Ahmad, N., Awan, M. J., Nobanee, H., Zain, A. M., Naseem, A., & Mahmoud, A. 2024. Customer personality analysis for churn prediction using hybrid ensemble models and class balancing techniques. *IEEE Access*, *12*, 1865–1879.

Ahn, J., Hwang, J., Kim, D., Choi, H., & Kang, S. 2020. A survey on churn analysis in various business domains. *IEEE Access*, *8*, 220816–220839.

Ahmed, U., Khan, A., Khan, S. H., Basit, A., Haq, I. U., & Lee, Y. S. 2019. Transfer learning and meta classification based deep churn prediction system for telecom industry. *arXiv*.

Bose, R. 2002. Customer Relationship management: key components for IT success. *Industrial Management and Data Systems*, *102*(2), 89–97.

Chandar, M., and Krishna, P. A. L. 2006. Modeling churn behavior of bank customers using predictive data mining techniques. Proc. Nat. Conf. Soft Comput. Techn. Eng. Appl. (SCT), pp. 24-26.

Dahiya, K., & Bhatia, S. 2015. Customer churn analysis in telecom industry. *ICRITO*.

De Andrade Moral, R., Chen, Z., Zhang, S., McClean, S., Palma, G. R., Allan, B., & Kegel, I. 2022. Profiling television watching behavior using Bayesian hierarchical joint models for Time-to-Event and Count data. *IEEE Access*, *10*, 113018–113027.

Deng, X., Oda, S., Kawano, Y., 2023. Graphene-based midinfrared photodetector with bull' s eye plasmonic antenna. *Optical Engineering*, 62(9), p. 097102-097102.

Ebiaredoh-Mienye, S. A., Esenogho, E., & Swart, T. G. 2021. Artificial neural network technique for improving prediction of credit card default: A stacked sparse autoencoder approach. *International Journal of Power Electronics and Drive Systems*, *11*(5), 4392.

Goodfellow, I., Bengio, Y. and Courville, A. 2016. Deep Learning, Cambridge. MA, USA:MIT Press.

Hassouna, M. S., Tarhini, A., Elyas, T., & AbouTrab, M. S. 2015. Customer churn in Mobile Markets: A comparison of Techniques. *International Business Research*, *8*(6).

Jahromi, A. T., Sepehri, M. M., Teimourpour, B., & Choobdar, S. 2010. Modeling customer churn in a non-contractual setting: the case of telecommunications service providers. *Journal of Strategic Marketing*, *18*(7), 587–598.

Komenar, M. 1996. Electronic marketing.

Lee, E., Jang, Y., Yoon, D., Jeon, J., Yang, S., Lee, S., Kim, D., Chen, P. P., Guitart, A., Bertens, P., Periáñez, Á., Hadiji, F., Müller, M., Joo, Y., Lee, J., Hwang, I., & Kim, K. J. 2019. Game data mining competition on churn prediction and survival analysis using commercial game log data. IEEE Transactions on Games, 11(3), 215–226.

Nath, S. V. and Behara, R. S. 2003. Customer churn analysis in the wireless industry: A data mining approach. Proc. Annu. Meeting Decis. Sci. Inst., vol. 561, pp. 505-510.

Panimalar, S. A., & Krishnakumar, A. 2023. Customer churn prediction model in cloud environment using DFE-WUNB: ANN deep feature extraction with Weight Updated Tuned Naïve Bayes classification with Block-Jacobi SVD dimensionality reduction. Engineering Applications of Artificial Intelligence, 126, 107015.

Parvatiyar, A. and Sheth, J. N. 2001. Customer relationship management: Emerging practice process and discipline. J. Econ. Social Res., vol. 3, no. 2.

Pamina, J., Raja, J., Bama, S. S., Soundarya, S., Sruthi, M. S., Kiruthika, S., Aiswaryadevi, V. J., & Priyanka, G. 2019. An effective classifier for predicting churn in telecommunication. Journal of Advanced Research in Dynamic and Control Systems, 11, 221–229.

Qiu, Y., Chen, P., Lin, Z., Yang, Y., Zeng, L., & Fan, Y. (2020, June). Clustering Analysis for Silent Telecom Customers Based on K-means++. In 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conf. (ITNEC) (Vol. 1, pp. 1023-1027). IEEE.

Seo, D., & Yoo, Y. 2023. Improving shopping mall revenue by Real-Time Customized digital coupon issuance. *IEEE Access*, *11*, 7924–7932.

Shaw, M. J., Subramaniam,C., Tan, G. W. and Welge, M. E. 2001. Knowledge management and data mining for marketing. Decis. Support Syst., vol. 31, no. 1, pp. 127-137.

Sugaya, T., Deng, X., 2019. Resonant frequency tuning of terahertz plasmonic structures based on solid immersion method. *2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves*, p.1-2.

Umayaparvathi, V., & Iyakutti, K. 2012. Applications of data mining techniques in telecom churn prediction. *International Journal of Computer Applications*, *42*(20), 5–9.

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. 2019. A Churn Prediction Model using Random Forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, *7*, 60134–60149.

Verbraken, T., Verbeke, W., & Baesens, B. 2014. Profit optimizing customer churn prediction with Bayesian network classifiers. *Intelligent Data Analysis (Print)*, *18*(1), 3–24.

Zhang, R., Li, W., Tan, W. M., & Mo, T. 2017. Deep and shallow model for insurance churn prediction service. *SCC*.