

# Integration of Churn Rate Prediction and Customer Lifetime Value Estimation using Temporal Convolution Networks

Suzana Isaac  
PG Student  
University of Birmingham

**Abstract**— Churn rate and Customer Lifetime Value (CLV) are an integral part of a Customer Relationship Management (CRM). Churn rate is calculation the percentage of the lost customers in a given total of customers at a certain time period. CLV is a measure to show the aggregate revenue an enterprise can expect from a single user version throughout the entire business relationship. To achieve the accurate predictions and estimation, using neural networks or machine learning techniques is crucial for the integration of CLV and Churn rate. By using online retail enterprise data which consists of half a million customers, data is collected to gather information regarding the historical transaction data of the customers. The primary aim of the research is to engineer this data by using a robust model to achieve state of the art results to enhance the CRM strategy. Leveraging Temporal Convolutional Networks (TCNs) to handle sequential data, it will capture long-term dependencies in customer transaction behavior. This research aims to endow the business models with deeper insights on customer behavior which maximizes the customer value with state-of-the-art results.

**Keywords**— Churn rate, Customer Lifetime Value (CLV), Temporal Convolution Networks (TCNs), Mean Average Error (MAE), Customer Relationship Management (CRM)

## I. INTRODUCTION

Achieving profitability and sustaining growth are primary requirements of any business model, may it be subscription services or online retail stores. Since it is a competitive landscape to retain the customers and amplify their lifetime value, Churn rate and Customer Lifetime Value (CLV) estimation prediction are vital key analytical tasks. Churn rate is calculation the percentage of the lost customers in a given total of customers at a certain time period. CLV is a measure to show the aggregate revenue an enterprise can expect from a single user version throughout the entire business relationship. For example, if a gym has accrued thousand members in the first week of the month and lose 50 members at the end of the month then the churn rate is 5%. CLV is when the member pays ten dollars per month and lasts for two years then total revenue from that member is \$240. The project addresses how to accurately predict the customer churn and estimate CLV by using the online retail enterprise data which consists of half a million customers. By leveraging Temporal Convolution Networks (TCN), this research aims to provide a robust result that can provide actionable insights to businesses effectively by using the sequential customer purchase data. The key issues to be tackled

are as follows: predictions of customers that are likely to churn based on the historical transactional data and usage patterns, estimation of long term fiscal valued customers for the company considering their historical transactions and integrating both Churn rate and CLV by using a single model to help highlight retention efforts for businesses. Major objective is to increase the model accuracy and optimize the CRM. Focusing on these key issues, this research will maximize the customer value and provide deeper insights of the purchase history which will give the ability to make it a profitable organization.

## II. LITERATURE REVIEW

Prediction of Churn rate and Estimation of CLV has been proposed with many ML techniques. The information on customer churn prediction and Customer Lifetime Value (CLV) estimation features several methods and models, comprising machine learning techniques like Random Forest, Support Vector Machine (SVM), and neural network. The papers offer insights into these methods and their applications.

The analysis by Nisha Gurung, Md Rokibul Hasan, Md Sumon Gazi, and Faiaz Rahat Chowdhury shows the efficiency of Random Forest and Decision Tree models in customer churn prediction. These models can help establishments recognize the causes leading to churn and progress aimed retention strategies, contributing to sustained revenue and economic growth. With an accuracy of 96.25% for Random Forest, they have achieved the churn rate suitable for many organizations [1]. Then there is a study of CLV in fast management strategy standpoint which significantly increases the profits of insurers. This was authored by Mathias Valla, Xavier Milhaud, Anani Olympio where they used tree-based survival model like Random Survival Forest (RSF), and Gradient Boosting Survival Model (GBSM) to improve the prediction and targeting strategies [2]. These papers show high accuracy but combining the strategy using Tree models will be more complex and are prone to overfitting the model.

The research on bank customers to predict Churn rate conducted by authors Pahul Preet Singh, Fahim Islam Anik, Rahul Senapati, Arnav Sinha, Nazmus Sakib, Eklas Hossain using SVM. They used this algorithm which finds the distance of the border of the required churn. For the experiment an accuracy of 83% was achieved [5]. There is a study the examines using SVM for CLV prediction and is used for customer profiling in digital start up where they clustered customers of different types and

used SVM model to accurately estimate customer behavior and sales potential [6]. With Support vector Regression (SVR), using longitudinal data using MAE there was a good result with a retention rate of 46.59% [7]. SVM has the advantage in handling nonlinear relationships with imbalanced data which is essential for CLV and churn rate. SVM is a great tool for customer profiling and segmentation. This shows that models that have been proposed so far do not essentially take in the sequential nature of customer interactions.

Using TCN is also highly accurate for CLV shown in research conducted by Zhiyuan Zhou, Li Lin, Hai Wang, Xiaolei Zhou, Gong Wei, and Shuai Wang. They experimented on real world dataset from JD.com which is a major supply chain platform. TCN has outperformed using traditional ML models. They achieved a normalized mean average error (MAE) of 0.3378 in CLV prediction, which is an improvement of 16.3% over baseline models [3]. Churn prediction where data collected from WSDM-KKBox dataset have received an accuracy of 60% using TCN [4]. TCNs can be incorporated to solving sequential data more effectively that understands long range dependencies which is primal for CRM strategy. The gap to be filled is to integrate both the CRM strategies using a single model. This approach can fill the gap where traditional methods dominate but may not fully capture the complexity of sequential patterns in customer behavior.

Table 1. Comparison of the Churn rate and CLV

Accuracy and Retention rate	Random Forest	SVM	TCN
Churn Accuracy	96.25%	83%	60%
CLV Retention Gain	23.90%	Not mentioned With SVR 46.59%	16.3%

### III. DATA AND METHOD

The Online Retail Dataset sourced from Kaggle contains transactional data of an online retail store based in the UK. This dataset is often used for analyzing customer purchasing behavior, predicting customer churn, and estimating Customer Lifetime Value (CLV). Using this dataset consists of half a million customer invoices (to be exact 541910 rows). The columns in the csv file include Invoice Number, Stock Code, Description, Quantity, Invoice Date, Unit Price Customer ID and Country of Purchase. The data spans over a specific period, capturing transactions from December 2010 to December 2011. The methodology involves data collection, preprocessing, model development, and deployment in a systematic manner. Cleaning of the data involves handling missing values and removing duplicate values. Then feature engineering of the dataset will be conducted to be ready as a model input. The effective use of TCNs can model the sequential data that captures long term dependencies and use of regularisation to prevent overfitting.

### IV. MODEL ARCHITECTURE

Figure 1 shows the visual representation of the model of integration of prediction of Churn rate and estimation of CLV using TCNs. Input layer receives sequences of monthly aggregated features of the customer. These features include MonthlyTotalPrice, MonthlyTransactions, AvgOrderValue, and DaysSinceLastPurchase. For example, if the sequence length is 6 months, the input for each customer would be represented as a 6x4 matrix, where 6 denotes the months and 4 denotes the features. MonthlyTotalPrice represents the total transaction amount for each month, MonthlyTransactions indicates the number of transactions for each month, AvgOrderValue provides the average order value for each month, and DaysSinceLastPurchase measures the last purchase number of days. This structured input layer ensures that the model captures comprehensive information about customer behavior over time.

TCNs are specifically designed to handle sequential data, making them ideal for tasks that involve time-series analysis. The TCN architecture (Figure 2) comprises multiple dilated convolutional layers, each with increasing dilation rates. These dilated convolutions enable the network to have a wider receptive field without significantly increasing the number of parameters, allowing the model to learn dependencies over various time scales [8]. Additionally, residual connections are incorporated within the TCN layers. These connections skip one or more layers, facilitating the flow of gradients through the network. This approach addresses the vanishing gradient problem and enables the training of deeper networks, ensuring that the model can learn complex patterns effectively. The output is split into two distinct paths: one for churn prediction and one for CLV estimation after passing the TCN layers. Churn Prediction output includes a fully connected layer that processes the output from the last TCN layer. A sigmoid activation function is then applied to convert the output into a probability score ranging between 0 and 1, indicating the likelihood of churn. The final output is the churn probability. CLV Estimation output is similar to the churn prediction path, this path includes a fully connected layer that processes the output from the last TCN layer. However, instead of a sigmoid activation, a linear activation function is used to convert the output into a continuous value representing the estimated CLV. The final output is the estimated CLV.

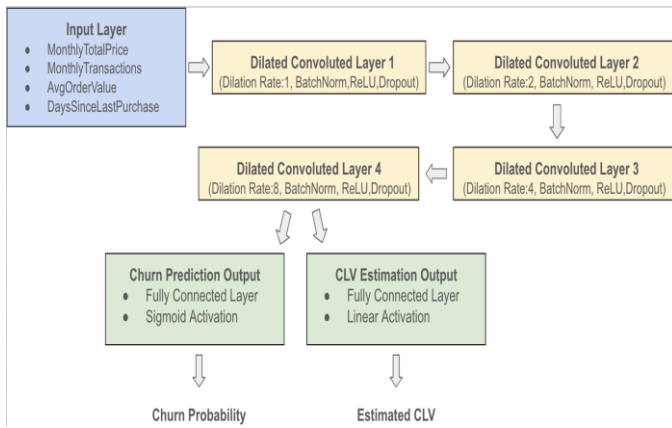


Figure 1. Workflow Diagram of the Model

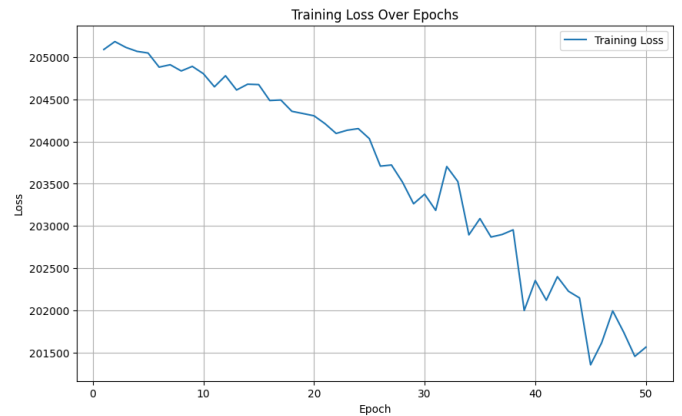


Figure 3. Training Loss of Churn and CLV Over Epochs

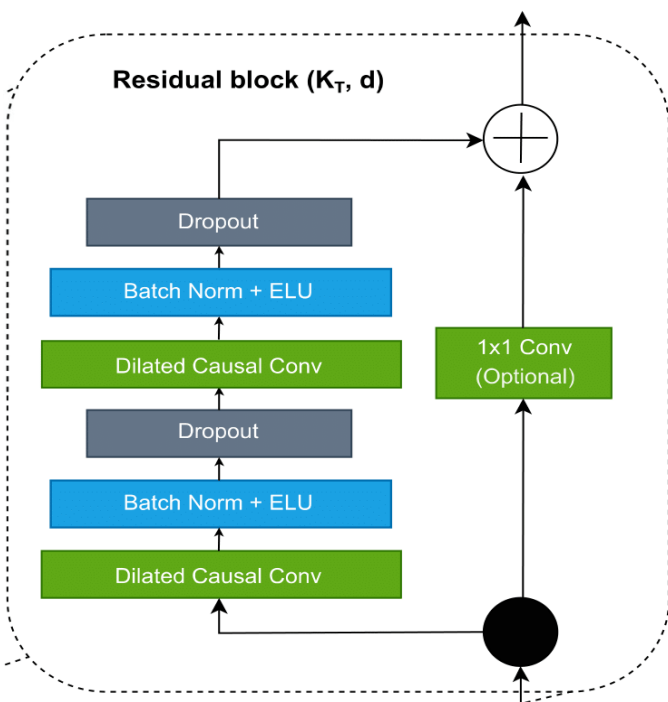


Figure 2. TCN Architecture

### V. RESULT

Figure 3 shows the loss values are decreasing over epochs, indicating that the model is learning and improving its performance over time. The test loss shown in figure 4 for churn prediction (0.54) and the accuracy (81%) suggest that the model performs with state-of-the-art results in predicting churn using TCN. The MAE of 86.28 shows the average error in the CLV values. This shows that the retention rate is 13.72%. The MAE represents the average absolute difference between the predicted and actual CLV values. This shows that combining the prediction of Churn and estimation of CLV gives better results as compared to the previous models. The reason as to showing MAE than accuracy for CLV is that it is an estimation and not a singular value. Integration of both techniques in a singular model progressed the result.

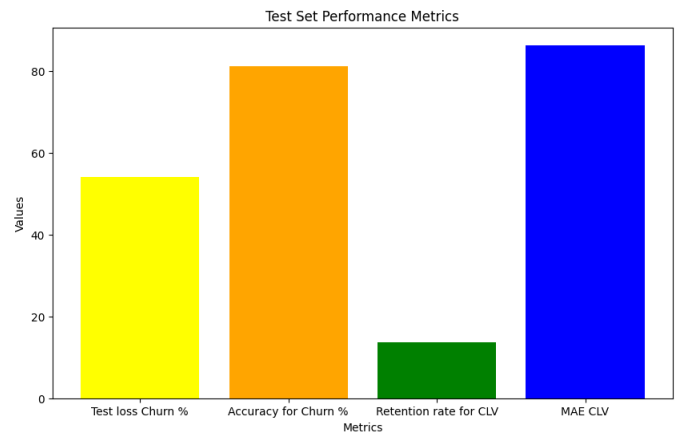


Figure 4. Test Set Performance Metrics

### VI. CONCLUSION

Applying TCNs for CLV and churn prediction can potentially lead to more accurate and efficient models by leveraging the strengths of deep learning for sequential data. This novel approach using TCNs integrated to Churn rate and CLV is expected to outperform traditional methods in handling the temporal aspects of customer data, leading to better predictions of churn and more accurate CLV estimates. Unlike RNNs and LSTMs, TCNs allow for parallel processing, making training faster and more efficient. TCNs maintain the input sequence length through causal convolutions, ensuring that there is no information leakage from future time steps. The main contribution is the effective modeling of sequential dependencies, which are crucial in understanding customer behavior over time.

By focusing on these strengths, the research can significantly advance the current methodologies for customer churn predictions and CLV estimation, providing a more powerful tool for businesses to manage customer relationships and improve retention strategies.

## REFERENCES

- [1] Nisha Gurung et al. (2024) 'AI-based customer churn prediction model for business markets in the USA: Exploring the use of AI and Machine Learning Technologies in preventing customer churn', *Journal of Computer Science and Technology Studies*, 6(2), pp. 19–29. doi:10.32996/jcsts.2024.6.2.3x.J.
- [2] I Valla, M., Milhaud, X. and Olympio, A. (2023) 'Including individual customer lifetime value and competing risks in tree-based lapse management strategies', *European Actuarial Journal*, 14(1), pp. 99–144. doi:10.1007/s13385-023-00358-0.
- [3] K Zhiyuan Zhou, Li Lin, Hai Wang, Xiaolei Zhou, Gong Wei, and Shuai Wang. (2024) 'A Cross Domain Method for Customer Lifetime Value Prediction in Supply Chain Platform' In *Proceedings of the ACM on Web Conference 2024 (WWW '24)*. Association for Computing Machinery, New York, NY, USA, 4037–4046. <https://doi.org/10.1145/3589334.3645391>
- [4] Subramanian, R.S. et al. (2024) 'Ensemble-based Deep Learning techniques for customer churn prediction model', *Kybernetes* [In press]. doi:10.1108/k-08-2023-1516.
- [5] Singh, P.P. et al. (2024) 'Investigating customer churn in banking: A machine learning approach and visualization app for data science and Management', *Data Science and Management*, 7(1), pp. 7–16. doi:10.1016/j.dsm.2023.09.002.
- [6] Kasem, M.S., Hamada, M. and Taj-Eddin, I. (2023) 'Customer profiling, segmentation, and sales prediction using AI in direct marketing', *Neural Computing and Applications*, 36(9), pp. 4995–5005. doi:10.1007/s00521-023-09339-6.
- [7] Chen, Z. Y., & Fan, Z. P. (2013). Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach. *Knowledge-based Systems*, 43, 123–134. <https://doi.org/10.1016/j.knosys.2013.01.022>
- [8] Ingolfsson, T. M., Hersche, M., Wang, X., Kobayashi, N., Cavigelli, L., & Benini, L. (2020). EEG-TCNet: An Accurate Temporal Convolutional Network for Embedded Motor-Imagery Brain–Machine Interfaces. <https://doi.org/10.1109/smc42975.2020.9283028>