Predicting Customer Churn Using Deep Learning Neural Network

Austin Campbell [*12] **Christopher Sáez** [*12] **Wisdom Okwen** [*12] **Paul Signorelli** [*12] **Ethan Byrd** [*12]
Will Scurria [*12]

1. Introduction

Customer churn, the phenomenon of customers leaving a service or discontinuing their relationship with a business, presents a significant challenge to companies in highly competitive markets. Predicting customer churn is a crucial step toward implementing retention strategies and improving customer satisfaction. In the context of banking, customer churn can lead to substantial financial losses, particularly when valuable or long-term customers are lost. Leveraging modern deep learning techniques to predict churn accurately can empower organizations to take proactive measures aimed at retaining their customer base.

2. Motivation

The motivation for this project stems from the increasing need for businesses to understand and mitigate customer attrition. In the banking sector, where customer retention directly impacts profitability and growth, predicting churn with high accuracy can provide a competitive advantage. By identifying potential churners early, banks can design targeted interventions to improve customer experience, loyalty, and retention. This project aims to build a robust deep learning model that can accurately predict customer churn, offering valuable insights to stakeholders and enhancing decision-making processes.

3. Methodology

The methodology for this project involved collecting a dataset derived from real bank data from 2022, which was slightly modified to preserve privacy and enhance analytical richness. The dataset comprises multiple features describing customer demographics, transaction history, account activity, and other relevant attributes. Preprocessing steps included handling missing data, normalizing numerical features, and encoding categorical variables to prepare the data for modeling.

To predict customer churn, we employed both traditional machine learning algorithms and deep learning architectures. Traditional machine learning models implemented include Logistic Regression, Random Forests, and Support Vector Machines. These models were evaluated for accuracy using standard performance metrics such as precision, recall, and F1-score.

For the deep learning approach, we constructed a neural network consisting of several fully connected layers. The architecture began with an input layer corresponding to the number of features in the preprocessed data, followed by hidden layers utilizing ReLU and linear activation functions. To improve the model's generalization capabilities and prevent overfitting, we applied dropout regularization after each hidden layer (Nitish Srivastava, 2014). Dropout randomly deactivates a fraction of neurons during training, forcing the network to learn more robust features.

Specifically, our neural network consists of five hidden layers, including a hidden linear trainable layer, ReLU activation functions after the first two linear layers, and a dropout rate of 0.3 applied after each ReLU function. The model was trained using the Adam optimizer, a variant of stochastic gradient descent known for its computational efficiency and low memory requirements (Diederik P. Kingma, 2015). The loss function employed was binary cross-entropy, appropriate for the binary classification task of predicting whether a customer would churn or not.

Training was conducted with 50 epochs and a batch size of 32. Early stopping was implemented to halt training when validation performance ceased improving, thus preventing overfitting. We also employed techniques such as learning rate scheduling to enhance convergence speed.

Comparative analysis between traditional machine learning models and the deep learning model demonstrated that the latter achieved superior predictive performance. However, the deep learning model required more computational resources and training time. The inclusion of dropout regularization significantly contributed to the model's robustness by mitigating overfitting.

Overall, our methodology underscores the utility of combining traditional machine learning models with modern deep learning techniques to achieve accurate and reliable predictions. The following sections will delve into the results

[*]Equal contribution [1]COMP 560 class project [2]University of North Carolina, Chapel Hill, United States.

obtained from these models and discuss their implications for customer churn prediction.

References

Diederik P. Kingma, J. L. B. Adam: A method for stochastic optimization. In *ICLR 2015*, 2015.

Nitish Srivastava, e. a. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research 15*, 2014.

A. You *can* have an appendix here.

You can have as much text here as you want. The main body must be at most 8 pages long. For the final version, one more page can be added. If you want, you can use an appendix like this one.

The `\onecolumn` command above can be kept in place if you prefer a one-column appendix, or can be removed if you prefer a two-column appendix. Apart from this possible change, the style (font size, spacing, margins, page numbering, etc.) should be kept the same as the main body.