**Research Methodology**

**Assignment-1**

Kavinesh K

CB.SC.P2CSE24023

# Explainable Customer Churn Prediction Model Based on Deep Learning

**Authors:**

- Shi yang, Li
- Ze lin, Shen

## Why Customer Churn Matters for Businesses

Churn prediction involves identifying at-risk customers who are likely to cancel their subscriptions or abandon their accounts.A churn model works by passing previous customer data through deep learning models to identify the connections between features and targets and make predictions about new customers.

A common example is people cancelling Spotify/Netflix subscriptions. So, Churn Prediction is essentially predicting which clients are most likely to cancel a subscription i.e 'leave a company' based on their usage of the service.

For example, if you got 1000 customers and lost 50 last month, then your monthly churn rate is 5 percent. Predicting customer churn is a challenging but extremely important business problem especially in industries where the cost of customer acquisition (CAC) is high such as technology, telecom, finance, etc.However, most existing prediction models, especially those using deep learning, are hard to understand and interpret. This makes it challenging for companies to figure out why customers are leaving and what they can do to prevent it.

## A Better Way to Predict Customer Churn

To tackle this issue, the authors suggest a new customer churn prediction model that is both accurate and easy to interpret. They use a deep learning approach with the TabNet model, which is designed to handle data in tables and make its decision-making process more transparent. The model also uses resampling techniques to balance the data, especially since there are usually fewer customers who leave than those who stay.

## How the Model Works

The proposed method involves a few key steps:

- **Data Preprocessing**: The customer data is cleaned and prepared, removing unnecessary details, handling missing data, and converting categories into a machine-readable format.
- **Resampling Techniques**: To deal with the imbalance between churned and non-churned customers, three methods—SMOTE oversampling, SMOTETomek hybrid sampling, and Tomek Links undersampling—are used to create more balanced datasets.
- **Model Training with TabNet**: The balanced datasets are then used to train the TabNet model. This model uses attention mechanisms to carefully select and transform features at each step, which helps it learn more complex patterns while remaining interpretable. All the outputs are combined to make the final prediction about whether a customer will churn.

## Testing the Model on Real Data

The authors tested their model on two datasets from the telecom and e-commerce sectors to see how well it worked. They compared their model's performance with other popular algorithms like AdaBoost, GBDT, XGBoost, CatBoost, and LGBM. The results showed that their model, called MultiS-TabNet, had better accuracy and F1 scores, which means it was more reliable and still easy to interpret.

## Final Thoughts and Future Directions

The paper concludes that their new model effectively predicts customer churn and is easy to understand, helping businesses figure out why customers leave and what can be done to prevent it. The model outperforms other models in predicting accuracy while still being transparent about its decision-making process. Future work could explore even better resampling methods and look into combining different models to make the predictions even more accurate.

## References

[1] Eva Ascarza, Oded Netzer, and Bruce G S. Hardie. 2018. Some Customers Would Rather Leave Without Saying Goodbye. Marketing Science 37, 1(January 2018), 54-77. https://doi.org/10.1287/mksc.2017.1057

[2] Fangyuan Li, and Guanghua Xu. 2022. AI-driven customer relationship management for sustainable enterprise performance. Sustainable Energy Technologies and Assessments 52, 8(August 2022), 102103. https://doi.org/10.1016/j.seta.2022. 102103

[3] Karmela Ljubičić, Andro Merćep, and Zvonko Kostanjčar. 2023. Journal of Computational Science 67(March 2023), 101940. https://doi.org/10.1016/j.jocs.2022. 101940

[4] Petra P Šimović, Claire Y T Chen, and Edward W. Sun. 2023. Classifying the Variety of Customers' Online Engagement for Churn Prediction with a MixedPenalty Logistic Regression. Comput Econ 61(January 2023), 451–485. https: //doi.org/10.1007/s10614-022-10275-1

[5] Sarkar M, and De Bruyn A. 2021. LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning. Journal of Interactive Marketing 53, 1(February 2021), 80-95. https://doi.org/10.1016/j.intmar. 2020.07.002