**Bank Customer Churn Prediction: A Machine Learning Approach**

**Background**

Customer churn is the act of a specific customer leaving his current firm to use the services of a rival organization. It is a prevalent situation and has grown to be a serious issue and is one of the main difficulties that many businesses across the world are having to deal with (Wen Z et al 2019). Since gaining new customers can cost up to five times as much as satisfying and keeping existing customers, numerous studies demonstrate that limiting customer turnover results in financial savings (Xiao et al 2015). Traditional banks are worried of customers moving to non-traditional alternatives as the World Retail Banking Report of 2019, states that in the next three years, 66.8% of existing banking clients plan to or have already utilized a bank account from a non-traditional business (big tech or fintech) (Capgemini 2019)

**Aim**

The focus of this project is to use machine learning algorithms to build a customer churn predictor application to help banks predict the likelihood of a customer leaving the bank.

**Key Techniques:** Data mining, machine learning algorithm, churn prediction

**Objectives**

1. To conduct a literature review on the use of different machine learning techniques for customer churn prediction.
2. To collect a significant amount of customer data to enable the analysis and prediction of churning possibilities, clean and preprocess the data to a format usable for analysis and prediction.
3. To perform a comprehensive analysis on the customer data to understand how factors such as demography, geography, participation, account balance, etc. can influence customers churning possibilities.
4. To conduct feature engineering, feature scaling, and feature selection to make the data ready for training using machine learning algorithms.
5. To develop and train different machine learning models on the customer dataset to make a forecast or prediction of the customer's likelihood to quit the company.
6. To perform models’ evaluation and assessment using metrics such as accuracy, precision, and recall and choose the best-performing model for deployment.
7. To use the trained model to build a machine learning web application called "ChurnSage" to provide an interface for utilizing the trained model for prediction.

**Literature Review**

The article published by Lemos et al. (2022) evaluates supervised classifiers such as decision trees, logistic regression, k-nearest neighbours, elastic nets, SVMs, and random forests used in banking to predict customer churn using a unique dataset from a large Brazilian bank. The random forest model outperformed other models with a recall of 80.2% and a 1-specificity of 14.8%, suggesting it could be used in the banking environment to promote customer retention and maintain lasting relationships. The study found that customer frequency, credit volume, and product possession had higher predictive power than transaction volumes. The study concludes that machine learning can help banks understand customer behaviour, enabling proactive actions to reverse potential customer churn and mitigate revenue losses. The only issue with this research is that it has a false negative rate of 19.8% and a false positive rate of 14.8%. Similar to Lemos et al. (2022), Rahman and Kumar (2022) experimented using supervised classifiers such as the KNN, SVM, Decision Tree, and Random Forest classifiers, and the experiment was conducted on the churn modelling dataset from Kaggle. The result of the experiment shows that the accuracy of the Random Forest model after oversampling is superior to other models. The downside to this project is the use of only accuracy as the evaluation metric; other metrics such as precision, recall, Roc, etc. should be explored as well. In the paper published by Jiang et al. (2020), they use Python to clean and condense real bank data, constructing three customer churn warning models (logistic regression, decision tree, and neural networks) and comparing them, achieving above 92% accuracy in predicting bank losses for each of the models. The only issue with this approach is that it is overly dependent on accuracy as the only evaluation metric. The paper, "A Survey on Customer Churn Prediction Using Machine Learning Techniques," performs a thorough analysis of the different techniques employed by researchers in predicting client attrition. The researcher concluded that fusing SVM with boosting algorithms can increase performance and accuracy. A combination of the above strategies must be employed, and good prediction models can be continuously achieved (Chandrakala and Kumar 2016).

**Summary**

Customer churn is a significant issue for businesses worldwide, with limiting turnover resulting in financial savings. This project aims to use machine learning algorithms to build a customer churn predictor application to help banks predict customer churn likelihood. The project aims to analyze customer data, train models, evaluate models, and use the trained models to build a web application called "ChurnSage" for prediction.

**References**

JIANG, W. et al., 2020. On the build and application of bank customer churn warning model. *International Journal of Computational Science and Engineering*, 22(4), p. 404. Available from: http://dx.doi.org/10.1504/ijcse.2020.109400.

KUMAR, S. and D., C., 2016. A survey on customer churn prediction using machine learning techniques. *International Journal of Computer Applications*, 154(10), pp. 13–16. Available from: http://dx.doi.org/10.5120/ijca2016912237.

DE LIMA LEMOS, R.A., SILVA, T.C. and TABAK, B.M., 2022. Propension to customer churn in a financial institution: a machine learning approach. *Neural Computing & Applications*, 34(14), pp. 11751–11768. Available from: http://dx.doi.org/10.1007/s00521-022-07067-x.

RAHMAN, M. and KUMAR, V., 2020. Machine learning based customer churn prediction in banking. In: *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE. pp. 1196–1201.

SHARMA, A. and PANIGRAHI, D.P.K., 2013. A neural network based approach for predicting customer churn in cellular network services. *ArXiv [Cs.NE]*. Available from: http://arxiv.org/abs/1309.3945 [Accessed 1 Jun 2023].

WEN, Z. et al., 2019. Customer churn warning with machine learning. In: *Proceedings of the Fifth Euro-China Conference on Intelligent Data Analysis and Applications*. Cham: Springer International Publishing. pp. 343–350.

*World retail banking report 2019*, 2019. [online]. Capgemini. Available from: https://www.capgemini.com/news/press-releases/world-retail-banking-report-2019/ [Accessed 1 Jun 2023].