ANALYSING BANK CUSTOMER CHURN PREDICTION

D. Srikar[1], B Sathvika[2], Ch Ashritha[3], A Sainath[4], M Achuthanand[5],

[1] Assistant Professor, Dept, of CSE(Data Science), Sreyas Institute of Engineering and Technology, Nagole, Hyderabad
[2,3,4,5] UG Student, Dept of CSE-(Data Science), Sreyas Institute of Engineering and Technology, Nagole, Hyderabad

Abstract Customer retention is really a challenge in the financial industry because existing relationships are far cheaper than buying new ones. This report describes a churn prediction study using machine learning techniques in the banking industry to improve retention. Having a database with demographics, transaction history, and other account information allows banks to be on a positive footing with their customers. The ML algorithms used for the churn prediction included Logistic Regression, Random Forests, and Lavatory, where Random Forest Ensemble algorithms turned out to be the most stable and accurate compared to the simpler models. Since banks can quickly identify at-risk customers, estimate the expected churn rates over given periods of time, and design retention programs for each customer, it can help reduce churn and increase overall levels of customer satisfaction.

Keywords: Bank customer churn, machine learning, customer retention, Random Forest, logistic regression, churn rate.

1 Introduction

As we all know customer churn would be very helpful in future since banks can predict if a customer stays or leaves and can take necessary precautions. Main reason of customer churn includes poor customer management relationship. Today's customers are well educated and better informed about new and emerging approaches. That knowledge today has a major impact on behaviour affecting their purchase habits and because of this they have become more inclined toward treating the sale/product aspect of the world in a way that they call 'analysis-paralyse' to make better purchase decisions. Therefore, this presents a major difficulty for new generation service providers to think about customers that, create connection values. And corporations need to know and value their consumers. Liu and Shih [1] further develop the argument that consumers push organizations in an increasing competitive manner to innovate the introduction of marketing strategies and the need to meet expectations and aspirations to sustain loyalty and increase retention

2 Dataset

This study used data from Kaggle and simulates customer churn in the banking sector, Dataset which contains records for 64,000 bank customers with the aim of predicting whether or not the customer has exited (left) the bank, by case determining whether customer has either try to go away or he stays in the bank. The target variable is called Exited (binary data), with 1 designating the customer has exited the bank (is churned), and 0 indicating the account has (remains retained). According to the banks data, 51,137 were retained (positive class), and 12,863 customers have exited the bank (negative class). The data frame has 13 feature vectors which represent customer demographics, (e.g. Geography, Gender, Age), customer financial status (e.g. CreditScore, Balance, EstimatedSalary), and customer behaviours (Tenure, NumberOfProducts, HasCrCard, IsActiveMember), as well as other factors that can be treated as input predictors in churn modeling and analysis.

3 Proposed System

This research primarily aims to develop and implement an effective and reliable customer churn prediction tool for the banking market. Because of active competitiveness in the banking market, it does not need emphasize that retaining customers is key for retaining profitability, letting customers churn is one of the most important issues in retaining market position. To solve this problem, the system will use a set of advanced ML algorithms to process huge amounts of real-time data, the algorithms would then be programmed to recognize patterns and behaviours that occurred before a customer churned. The system would act to identify high risk customers early so the banks could take proactive intervention. For example, once the model could demonstrate that a customer was likely to churn, banks could then have implemented retention strategies to improved customer satisfaction , such as personalized customer care, better customer service, or rewards programs for the customer to build satisfaction and retain them which would also automated the bank's process of evolving and enhancing their customer retention policies which allows the bank not only to stop customer churn but also better their customer satisfaction and customer retention.

3.1 Implementation

3.1.1 Data Preprocessing

Data preprocessing is an important aspect of the data mining process. In fact, because they have a significant effect on anything that contributes to success in this process, is an important step. It has to cope with relevance, drift, and ambiguity of information. Any information or features that do not pertain to the specific subject of focus are irrelevant. This won't have any influence on the results of the classifiers. The forecast does not provide any knowledge about the weather, other than customer ID, last name, and location. These characteristics are not accidental omissions in this research.

3.1.2 Classification

To reduce errors many decision trees ae combined together Each tree is trained on a different subset of the data, and the outcome is the. Prediction is based on the mean (for regression) or majority vote (for classification). Each of the trees. Random forest comes from the fact that it is less likely to overfit than a single decision tree. Random forest is effective for exploring complex datasets as it requires little feature selection. Random forest is a machine learning method is a way to use multiple decision trees together to create a single complete model. With regression accuracy, decision trees amount the output of individual decision trees together to increase prediction accuracy and direct decision making. Overfitting: does both classification and regression. In contexts of customer attrition .Random forests can predict whether a customer will possibly leave, based on characteristics like payment records, age, gender, location etc.

3.1.3 Oversampling

In data processing, different methods are used to change the shape of dataset. Distribution of Classes in the Given Data. Given the vast difference in the dataset, with 51,137 positive class samples and 12,863 negative class samples, this is needed. However, since our sample of existing available data is limited, this research will opt for the following. Oversample the data. If you want to undersample, the data will be limited.

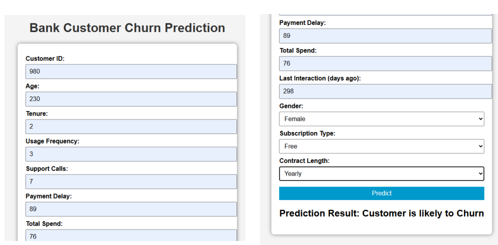
3.1.2 Selecting necessary features

In ML, selecting relevant features for analysis is called feature selection. The last stage of the output is at the highest quality. Not only does it reduce training time, it will help you fight the curse of high-dimensionality. Reduce the model's complexity.

4 Results and Discussion

In the banking World, a key area of focus for has emerged as customer engagement. Organizations understand, that keeping a loyal customer base is vital for long term success. One of the major threats that banks have to deal with is when customers churn (customers leaving). The ability to facilitate internal interventions for retention, coupled with how effective those interventions were are well assessable with logistic regression and random forests (rf). What is clear, is that in order to correctly adjudicate churn means retaining powerful predictive models. However, this undertaking is complicated because varying units in the banking World use some common definitions for gauging churn, however the variables they use such as transaction history, decorated account activity, service usage, and customer demographics are notably varied. The difficultly is creating a generic model for the entire banking industry, that can prognosticate churn primarily through established and more recent sources of data, which are completely unrelated to specific providers of service.

Fig 4.1 Output of the bank customer churn prediction



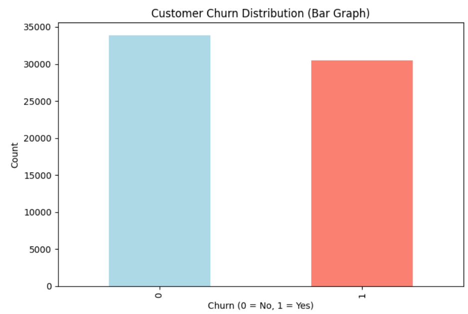


Fig. 4.2 Customer Churn Prediction Bar graph

5 Conclusion and Future directions

To conclude, the results from this project show that while random forests are appropriate for Customer churning need to be applied cautiously in the banking sector. Data preprocessing, in particular, unbalanced datasets must be addressed. This research demonstrated that churning prediction systems need to be adjustable to accommodate different input data, while also producing usable Indicator On time that allows banks to act pre-emptively and keep customers Avoiding Churning. The performed models were tested in numerous ways to observe how well they predicted customer churning. Therefore, the Off, random forest classifier with oversampling, would demonstrate. Random forests performed well due to their capability to address unbalanced data and determine complex relationships between variable**.**

6 References

[1] Manzoor, A., Qureshi, M.A., Kidney, E. and Longo, L., 2024. A review on machine learning methods for customer churn prediction and recommendations for business practitioners. *IEEE access*..

[2] Chayjan, M.R., Bagheri, T., Kianian, A. and Someh, N.G., 2020. Using data mining for prediction of retail banking customer's churn behaviour. *International Journal of Electronic Banking*, *2*(4), pp.303-320.

[3] Morik, Katharina, and Hanna Köpcke. "Analysing customer churn in insurance data–a case study." *Knowledge Discovery in Databases: PKDD 2004: 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pisa, Italy, September 20-24, 2004. Proceedings 8*. Springer Berlin Heidelberg, 2004.

[4] Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2024). A review on machine learning methods for customer churn prediction and recommendations for business practitioners. *IEEE access*.

[5] A.K. Ahmed, A. Jafar, K. Customer brainstorming in Aljouma, Telecom Machine Learning in Big Data Platform, Journal of Big Data 6 (1) (2019) 1-24.