

Customer Churn Prediction for Indian Bank

Problem Statement

Over the past few quarters, Indian Bank has observed a significant number of customers closing their accounts or switching to competitor banks. This trend has caused a substantial dent in the quarterly revenues and might drastically affect the annual revenues for the ongoing financial year, leading to a plunge in stocks and a reduction in market capitalization.

Objective

The primary objective of this project is to build a predictive model that can identify customers who are likely to churn in the near future with reasonable accuracy. An added bonus would be the model's ability to accurately estimate when the churn is likely to occur.

Definition of Churn

In the context of this problem, a customer is said to have churned if they have closed all their active accounts with the bank. However, the definition of churn can vary based on the context of the problem and business requirements. For instance, a customer not transacting for 6 months or 1 year can also be defined as having churned.

Metrics and Goals

Data Science-Related Metrics

The performance of the predictive model will be evaluated using the following metrics:

- **Recall:** This is the ratio of the total number of correctly predicted positive observations to the total actual positives. It is defined as

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

where TP is True Positive and FN is False Negative. The target for Recall is set to be greater than 70%.

- **Precision:** This is the ratio of correctly predicted positive observations to the total predicted positives. It is defined as

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

where FP is False Positive. The target for Precision is set to be greater than 70%.

- **F1-Score:** This is the harmonic mean of Precision and Recall and provides a better measure of the incorrectly classified cases than the Accuracy Metric. The target for F1-Score is set to be greater than 70%.

Business Metrics

While the data science metrics are usually top-down, a good practice is to consider the business metric to make at least half the impact of the data science metric. For example, if we take the Recall target as 70%, which means correctly identifying 70% of customers who are going to churn in the near future, we can expect that due to business intervention (offers, getting in touch with customers, etc.), 50% of the customers can be saved from being churned. This means at least a 35% improvement in the Churn Rate.

Conclusion

This project aims to leverage data science to address a critical business problem. The predictive model developed in this project will not only help Indian Bank retain its customers but also significantly improve its revenues. The model's performance will be continuously monitored and updated as necessary to ensure its effectiveness.