CUSTOMER CHURN PREDICTION IN THE BANKING INDUSTRY USING MACHINE LEARNING

Chapter 1: INTRODUCTION

* Overview

Customer churn has been a critical issue affecting the banking industry, impacting revenue and customer lifetime value. This project creates a predictive model that can be used to predict which customers are at risk of leaving allowing banks to take proactive actions to retain them.

* Problem Statement

To be able to identify customers that are about to leave requires analyzing multiple customer-related factors and developing an accurate predictive machine learning model.

* Objective

This project aims at developing a machine learning model that predicts customer churn for any bank and provide usable insights to help reduce churn.

Chapter 2: LITERATURE REVIEW

* Customer Churn Prediction in Banking Sector

Previous researches have shown that Customer churn in banking industry can be caused by various factors such as Account balance, Customer’s tenure, Age and Product usage.

* Challenges

Developing a predictive model to predict churn can be complicated by some factors such as class imbalance, where we see majority of customers don’t churn, this can lead to skewed models if not addressed properly

Chapter 3: METHODOLOGY

* Data Collection and Description

The dataset used was collected from Kaggle and it contains information about customer demographics, product usage, tenure, salary, account balances and many more.

* Data Preprocessing

The data was inspected by checking if it had missing values, duplicates and the datatypes of each feature. The data was unbalanced since it contained a high number of customers who churned than those who didn’t churn.

* Model Development

Here, I trained multiple machine learning models such as Logistic Regression and Random Forest to predict churn. The Random Forest model gave 100% accuracy proving to be the most efficient model for this prediction.

* Evaluation

I used different metrics such as checking accuracy of the model both on the training data and testing data. Used also precision to and F1-score to assess model’s performance and chose the best model for deployment.

Chapter 5: DISCUSSION

* Insights

1. The customers who have less account balance are most likely to churn.
2. Also, the customers who have less estimated salary are also most like to leave compared to the customer with higher salaries.
3. Customers whose credit score is low are likely to churn than the rest who have got higher credit score.
4. The Customers who a consuming the most products in the bank seem to churn from previous study.
5. In terms of age, the customers who are above 60 years of age are not leaving their respective banks but those between 50 and 60 years are most likely to churn

Chapter 6: CONCLUSION AND RECOMMENDATIONS

* Conclusion

In this project, I managed to develop a Random Forest machine learning model and the model shown 100% accuracy proving to be very effective and accurate to predict customers who are about to churn. This model can be used by banks to counter the churn risk.

* Recommendations

Here are some of the strategies that banks can use to reduce churn: -

* Targeting high-risk customers with personalized offers.
* Can offer a balanced Credit score to all customers since the customers who feel like they have less Credit score seem to churn the most.
* Offer more interesting and improved offers to young customers. Offers that can attract them more to the bank.