

PROJECT PROPOSAL

GROUP-12

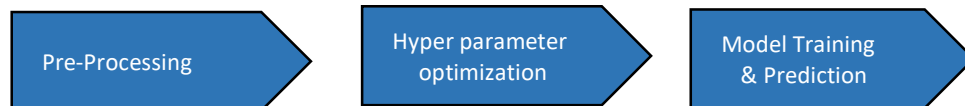
Title: Customer Churn Prediction for Bank in the U.S. Market:

Students: Varun Simha Reddy Thippireddy, Santhoshi Kareddy, Naveen Ajay Karasu, Venkata Sai Karthik Aleti, Gnana Naga Siva Sankar Kuralla

Abstract: Customer churn, the loss of customers to competitors, presents a significant challenge for banks. This project aims to understand customer churn behavior by employing various machine learning techniques on a bank customer dataset. By analyzing various factors like demographics, account activity, and product usage, we hope to build a model that accurately predicts which customers are most likely to leave the bank. This predictive capability can empower banks to implement targeted retention strategies, ultimately reducing churn and enhancing customer loyalty.

Problem: This project addresses the problem of classifying bank customers into two categories: those who will churn and those who will remain loyal. This is a binary classification problem, making machine learning algorithms like decision trees, Random Forest classifier, and neural networks suitable candidates. Our primary goal is to build a model with high accuracy, precision, and recall in predicting churn, allowing banks to identify at-risk customers proactively.

Data: We will utilize the provided bank customer churn dataset containing information on customer demographics, financial details, and churn status from U.S. banks. The dataset reportedly holds hundreds of examples with several relevant features. Data cleaning and preprocessing might be necessary before applying machine learning algorithms.



Methodology: Data Exploration: Analyze the provided dataset to understand its characteristics, distribution of features, and potential relationships with churn.

Model Selection and Training: Experiment with different machine learning algorithms like Decision trees, Random Forest, and potentially neural networks. Evaluate their performance metrics like accuracy, precision, recall, and F1 score and Confusion Matrix to select the best model for prediction.

Model Tuning and Optimization: Fine-tune the chosen model's hyperparameters to improve its performance on the dataset.

Evaluation and Comparison: Compare the performance of different models on a held-out test set. Consider baseline models like predicting all customers to stay or leave to gauge the significance of the chosen model's performance by drawing the ROC curve.

Interpretation and Insights: Analyze the final model's predictions and feature importance to understand the factors contributing most to customer churn.

Tools and Languages: We plan to leverage Python libraries like pandas, scikit-learn, and TensorFlow for data analysis, model development, and evaluation.

Expected Outcome: We aim to build a reliable and interpretable machine learning model for predicting bank customer churn. This model should assist banks in identifying customers at risk of leaving, enabling them to implement targeted retention strategies and improve customer loyalty. Additionally, the project will provide valuable insights into the characteristics and behaviors of churning customers, contributing to a deeper understanding of customer churn dynamics in the banking industry. This outline provides a comprehensive overview of your project, clearly defining the problem, approach, and expected outcome. Remember to adapt and refine it as you progress through your project, incorporating specific findings and results.

Datasetlink: <https://www.kaggle.com/datasets/shantanudhakadd/bank-customer-churn-prediction>