**Introduction**

This study aims to predict customer churn in the banking industry using machine learning algorithms. Customer churn, or the rate at which customers discontinue bank services, is a crucial factor in maximizing revenue and maintaining a positive reputation. The study utilizes a dataset from Kaggle consisting of 10,000 samples and 14 features, including age, gender, balance, tenure, and an "Exited" column indicating whether the customer has left the bank. The authors use exploratory data analysis to gain insights into the dataset, including the observation that 20% of loyal customers who have been with the bank for more than three years left. They also divide the data into training and testing sets and use supervised machine learning algorithms such as decision tree, extra gradient boosting, and random forest to build the predictive model. Finally, the authors evaluate the model's performance using accuracy scores and confusion matrices.

Previous research has shown the effectiveness of machine learning techniques in churn prediction and retention. The authors of this study use exploratory data analysis and feature engineering to improve the performance of the predictive model. The dataset used in the study includes customers from France, Germany, and Spain. The authors find that the number of customers who churned was highest in Germany, while France had the highest retention rate. They also find that male customers retained more than female customers, customers with credit cards churned more, and inactive members were more likely to churn.

The study's methodology involved using Python programming language for data analysis and machine learning algorithms. The authors used exploratory data analysis to gain insights into the dataset and feature engineering to improve the model's performance. They then divided the data into training and testing sets and used supervised machine learning algorithms to build the predictive model. Finally, they evaluated the model's performance using accuracy scores and confusion matrices.

In conclusion, the study demonstrates the effectiveness of machine learning algorithms in predicting customer churn in the banking industry. The authors use exploratory data analysis and feature engineering to improve the model's performance and evaluate its accuracy using standard metrics. The study provides insights into factors that affect customer churn and can help banks develop strategies to retain loyal customers.