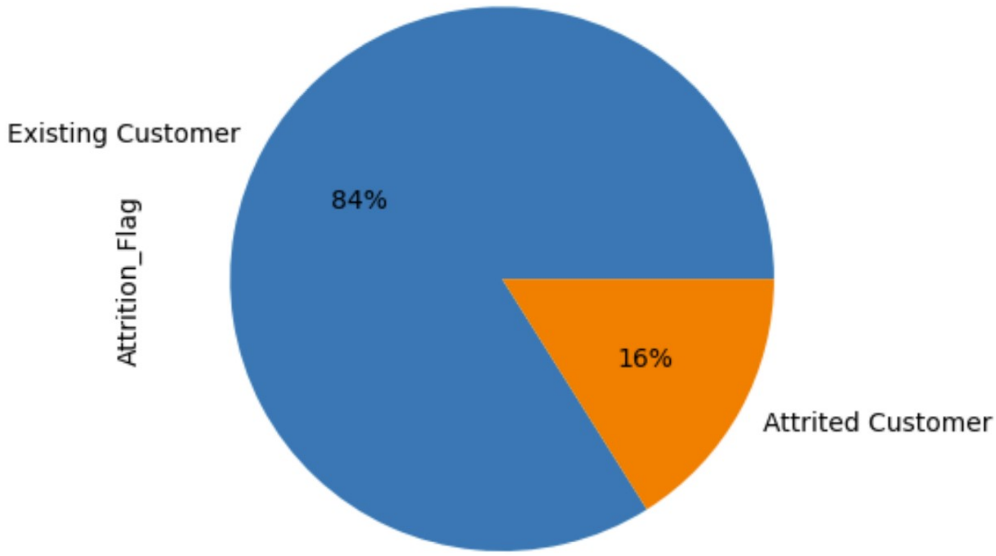
Introduction

One of the key challenges faced by banking institutions is understanding and addressing customer churn. The ability to predict which customers are likely to leave allows banks to proactively engage with them, provide better services, and potentially retain their valuable business. Our project stems from a real-world scenario at a bank where a growing number of customers are leaving their credit card services. The objective is clear: develop a model that can predict which customers are on the verge of churning. To achieve our goal, we will be processing on a dataset where we could know some basic characteristics of customer, who either still using their credit card, or have already out of the services.

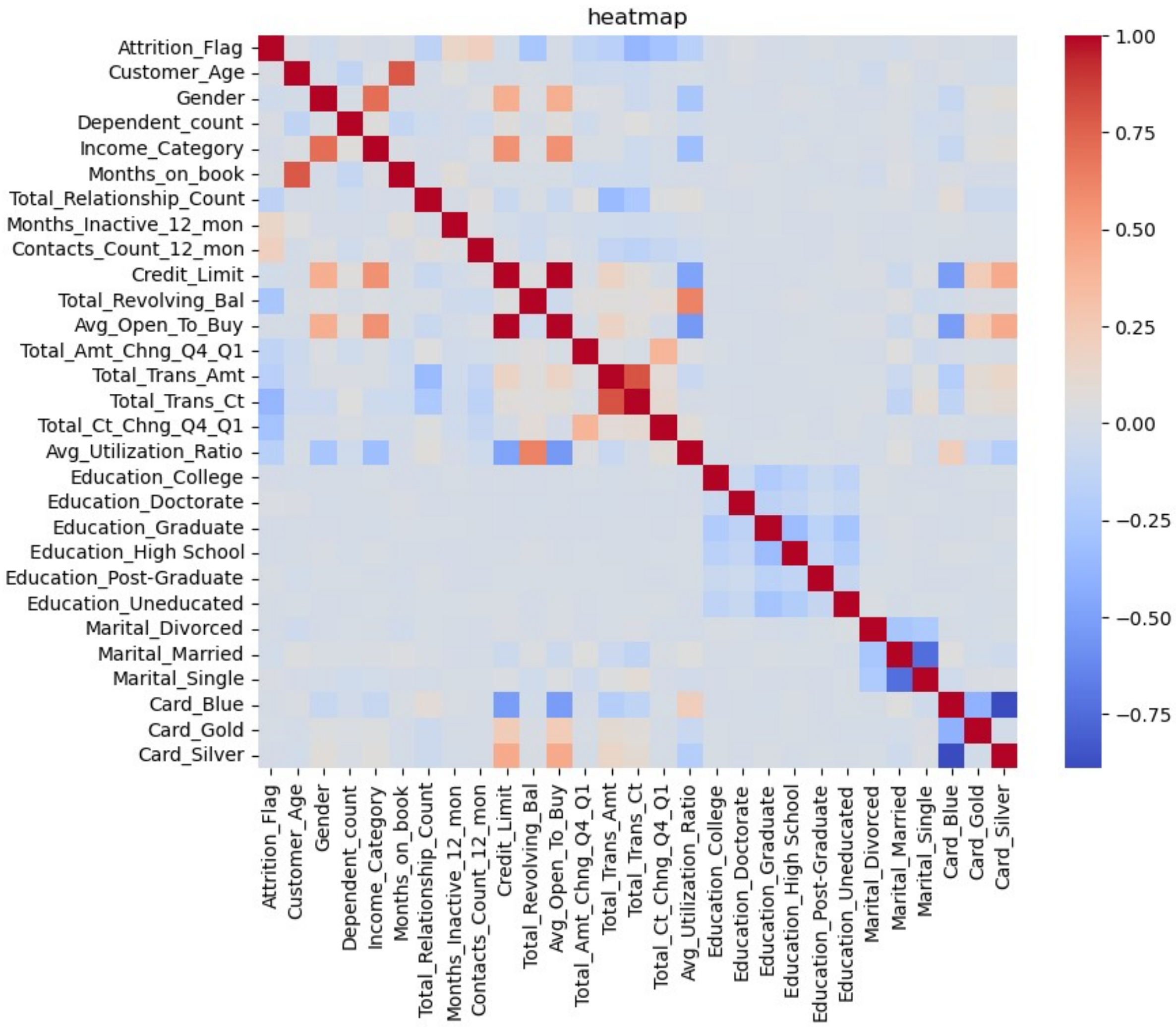
Data and Machine Learning Analysis

This report presents an analysis of customer churn in credit card services. Utilizing a dataset from Kaggle with 10,000 customer records and 23 features, the study aims to identify key factors influencing customer retention. Through machine learning techniques, including kNN, SVM, Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Stacking, the analysis provides insights into churn prediction and strategies to enhance customer loyalty.
Feature Engineering: The process involved data cleaning, rescaling, imputation, and analyzing inter-feature relationships using a heat map. For data visualization, we first need to understand the churn rate, which is the core data, and then compare different variables, such as the card holder's card tier and the cardholder's age.



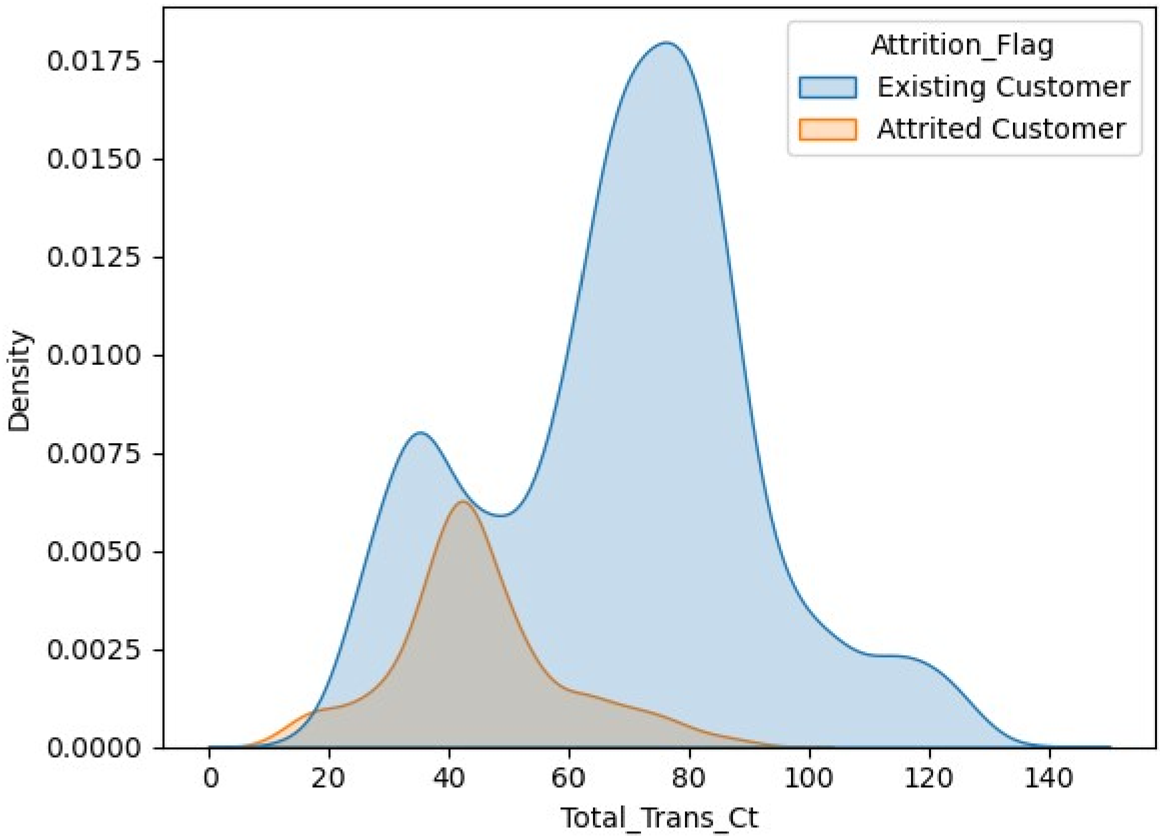
Machine Learning Models: Several models were employed. The kNN model offered simplicity but lacked precision in large datasets. SVM was effective for non-linear relationships but computationally intensive. Logistic Regression provided a probabilistic approach, ideal for binary outcomes. Decision Trees were intuitive but prone to overfitting. Random Forests improved this by ensemble learning. Gradient Boosting optimized on errors, and Stacking combined models for enhanced predictions.

Model Evaluation: Cross-validation and hyperparameter tuning were critical for model optimization. Performance metrics on unseen test data highlighted each model's effectiveness and limitations, emphasizing the importance of balancing accuracy with the risk of overfitting. the feature engineering process in their study on customer churn. It describes the use of binary encoding for variables like Attrition Flag and Gender, one-hot encoding for categorical features such as Card Category, and standardization for numerical features like Customer Age. The document also mentions the use of different numbers to represent Income Category levels. A heatmap was used to identify highly correlated features, leading to the dropping of some for improved accuracy. It also discusses unsuccessful attempts, like resampling to address imbalanced data, which decreased accuracy but increased recall, making it useful for identifying potential customer churn.



Conclusion

The analysis identifies significant factors contributing to customer churn in credit card services. While each model presented unique strengths, a combined approach using Stacking provided the most holistic understanding. The study underlines the necessity of continuous model evaluation and adaptation to emerging data trends. For the kernel density estimate (KDE) plot, which is used to visualize the distribution of a continuous variable. In this case, it represents the distribution of 'Total_Trans_Ct' (Total Transaction Count) for 'Existing Customer' and 'Attrited Customer'. The blue curve represents the density of the existing customers, and it has a peak at a higher transaction count, suggesting that most existing customers have a moderate to high number of transactions. The orange curve represents the density of the attrited customers, and it peaks at a lower transaction count, indicating that customers who have churned generally have fewer transactions.



Member	Proposal	Coding	Presentation	Report	
Zhixing Liu	1	1	1	1	
Ge Li	1	1	1	1	
Zifu Wang	1	1	1	1	
Jiapeng Wang	1	1	1	1	
Author Hu	1	1	1	1	