Propensity to Lapse Model building

STAMATOULA-GERASIMOULA MESOLORA

Registration number: f2822308

Course: Analytics Practicum II

July 2024

Problem

Anticipating customer churn within company's reward system. Despite successful operations and customer engagement through the collection and redemption of points, there is a need to accurately identify which customers are likely to churn, thereby optimizing retention strategies and improving overall business performance.

Objective

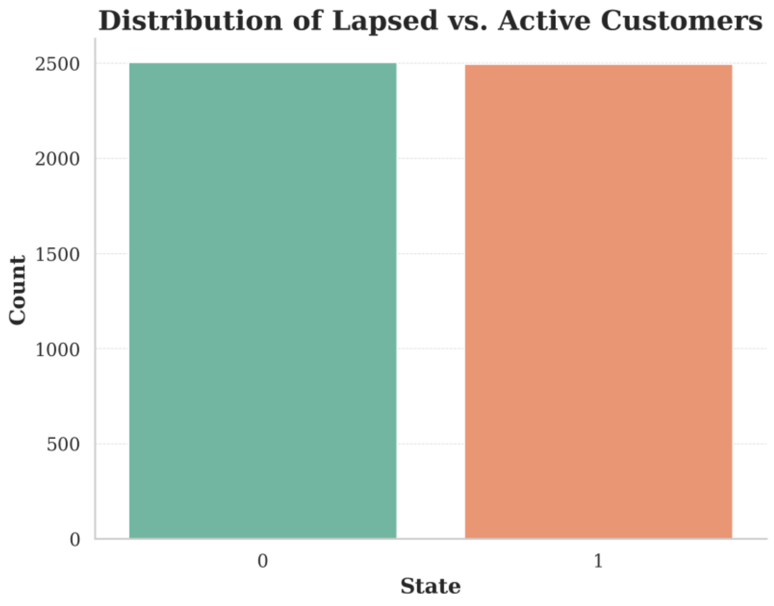
The objective is to develop a model which predicts customers' propensity to lapse utilizing their transaction data. The aim of the model is to analyze customer behaviors and identify accumulation patterns to accurately predict whether a customer will transition to a lapsed status or remain active.

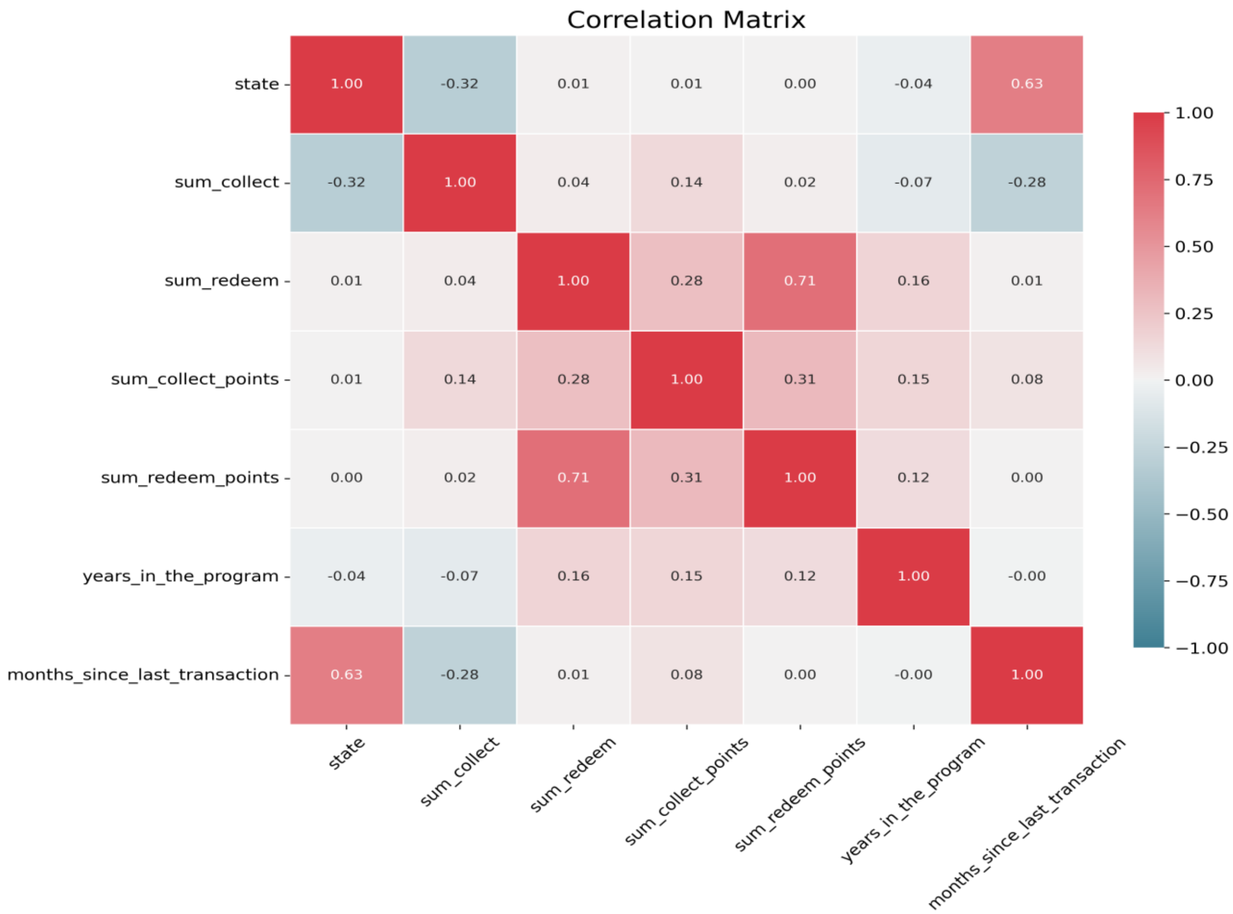
The given dataset consists of 5000 observations customers having the below fields :

- *State*: Lapsed status of the customer (Active, Lapsed).
- *Sum_collect*: Number of times a customer collected loyalty points.
- *Sum_redeem*: Number of times a customer redeemed loyalty points.
- *Sum_collect_points*: Total points collected by the customer.
- *Sum_redeem_points*: Total points redeemed by the customer.
- *Years_in_the_program*: Number of years since the customer's registration to the loyalty program.
- *Months_since_last_transaction*: Number of months passed since the customer's last activity.

Descriptive Statistics

- The below histogram visualizes the distribution of lapsed versus active customers within the initial dataset. The distribution of lapsed versus active customers within the initial dataset is equal.





- From the above Correlation Matrix, it seems that the number of points a customer has redeemed in total is highly correlated to the number of times a customer redeemed.

Model Construction

Models used for Prediction

- Logistic Regression: A statistical model that uses a logistic function to model the probability of a binary outcome. It is effective for capturing the linear relationship between the features and the log-odds of the target variable, allowing for straightforward interpretability and feature importance analysis.
- Decision Tree Classifier: A non-parametric model that splits the data into subsets based on feature values, forming a tree-like structure. It finds the optimal splits to classify the data by recursively partitioning it, which makes it powerful for capturing complex interactions and non-linear relationships between features.

Logistic Regression

For the variable selection of the Logistic Regression model, there has been executed a stepwise selection method. The selected variables for the Logistic Regression model are the following:

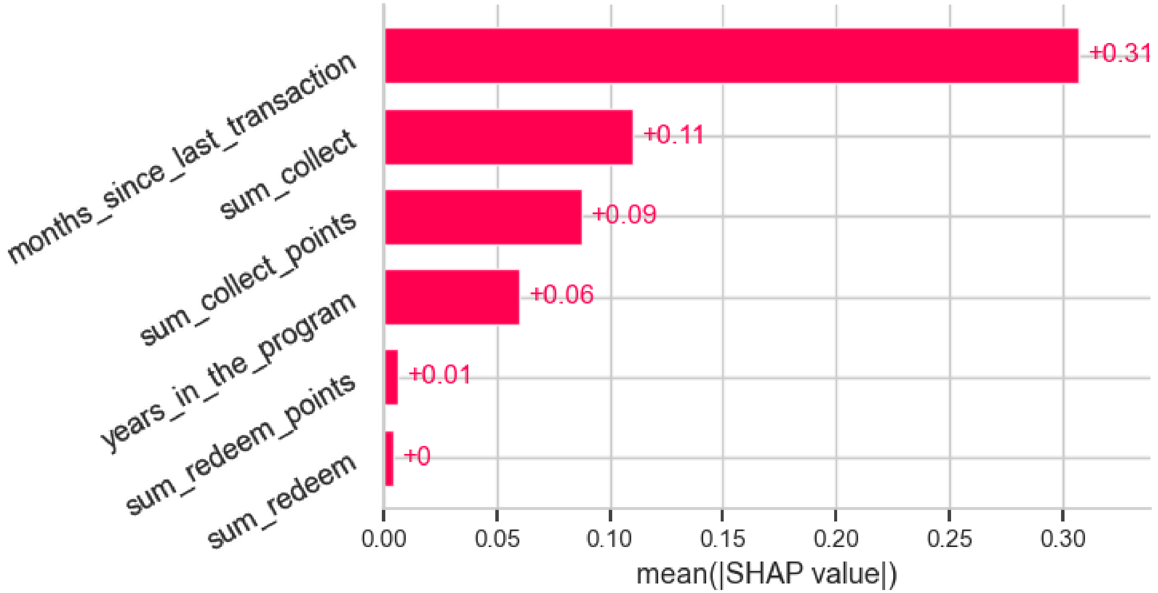
- Months since last transaction
- Sum of collections
- Years in the program

Decision Tree Classifier

For the Decision Tree model, we focused on the number of years since customer's registration to the program and divided this feature into three groups: less than 1 year, 1-5 years and more than 5 years.

We then calculated the average impact of each feature on customer churn for each group. This allowed the creation of a visual representation of the most important factors driving churn. The below bar chart reveals the top three variables that have the greatest influence on customer churn:

- ❖ Months since last transaction
- ❖ Sum of collections
- ❖ Sum of collection points



However, we have seen that the sum of collections is highly correlated to the sum of collection points, so **we are not going to include the sum of collection points in the model.**

Model Selection

The evaluation of has been done according to the below measures:

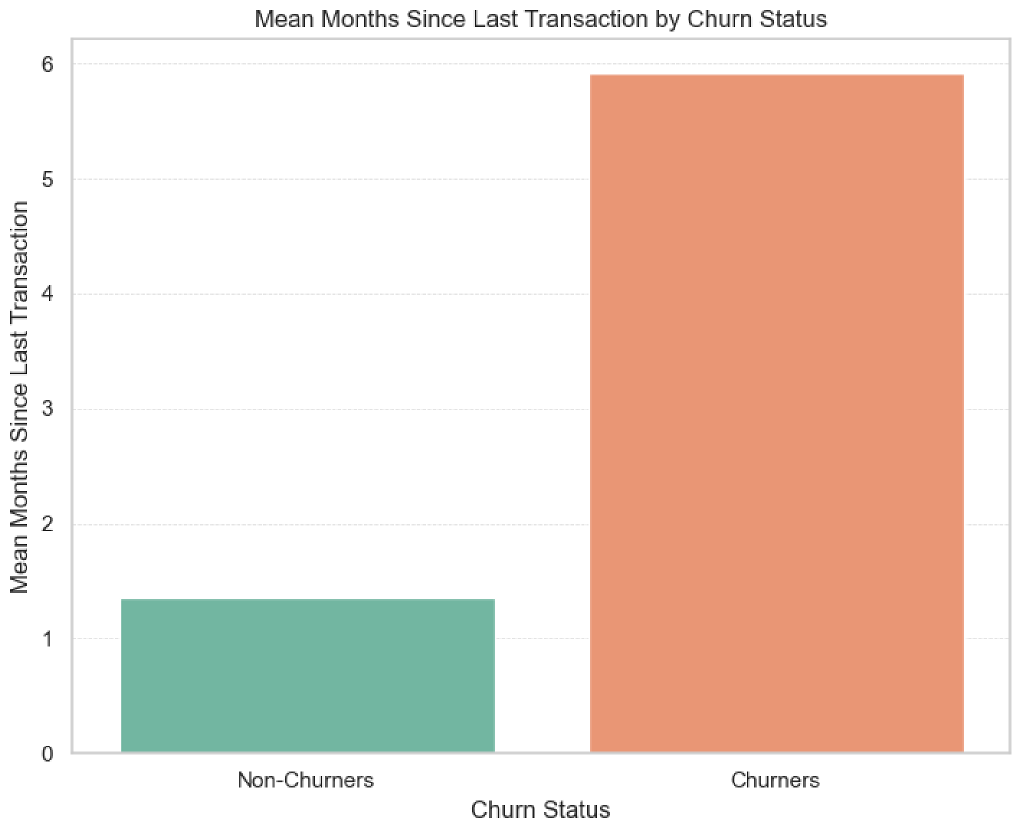
- Accuracy: The proportion of correctly classified instances out of the total instances.
- F1 Score: The harmonic mean of precision and recall, providing a balance between the two.
- Recall: The ratio of correctly predicted positive observations to all observations in the actual class.
- ROC AUC Score: The area under the receiver operating characteristic curve, representing the model's ability to discriminate between positive and negative classes.

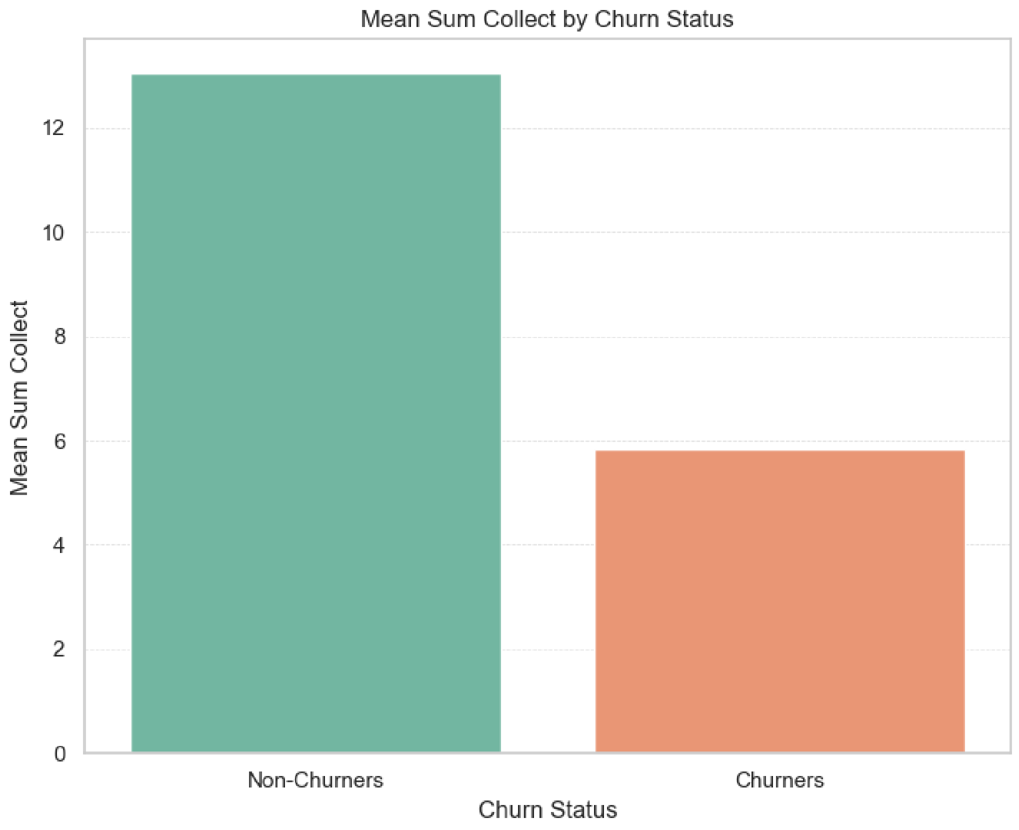
This comparative analysis helps in understanding the strengths and weaknesses of each model.

	Logistic Regression	Decision Tree
Accuracy	0.794	0.824
F1 Score	0.776	0.826
Recall	0.711	0.838
ROC AUC Score	0.793	0.824

The above table provides a clear comparison of the models' performance across the described metrics. It is evident that the **Decision Tree model** performs better than the Logistic Regression model.

Findings





- ❖ The average number of months since last transaction for Churners is significantly higher than for Non-Churners.
- ❖ Non-Churners transact every month and a half on average.
- ❖ Churners transact on average over a period of five months.

- ❖ The average number of times a customer collected is twice as much for non-churners than for churners.
- ❖ Non-Churners collect on average 12 times during the analysis period, while Churners collect approximately 6 times.

Recommendations

Enhance Engagement Strategies:

❖ Personalized Offers: Create targeting marketing campaigns aimed at customers who are predicted to lapse & Personalized incentives such as exclusive deals or discounts to encourage continued engagement.

❖ Regular Communication: Increase the frequency and relevance of communication with customers through newsletters, emails or app notifications that remind them of their points balance, offer special deals or provide updates on new ways to earn and redeem points.

Introduce Loyalty Program Enhancements:

❖ Tiered Loyalty Levels: Introduce tiered loyalty levels that provide increasing benefits and recognition as customers engage more with the program. This can create a sense of achievement and motivate customers to stay active.

❖ Bonus Points for Activity: Offer bonus points for consistent activity, such as a monthly collection or redemption. For example, reward customers with bonus points if they collect or redeem at least once every month.

Recommendations

Reactivate Dormant Customers:

- ❖ Win-back Campaigns: Design win-back campaigns targeted at customers who have shown signs of lapsing, such as exclusive offers for returning customers.
- ❖ Feedback Mechanism: Implement a feedback mechanism to understand why customers are lapsing and use this feedback to improve the loyalty program.

Optimize Communication Timing:

- ❖ Timing of Offers: Send reminders or special offers a month after customers' last transaction to encourage another interaction.

Improve Customer Experience:

- ❖ Expand Partnership Network: Expand the network of affiliated partners where customers can collect and redeem points. More options can increase the perceived value of the loyalty program and encourage more frequent participation.