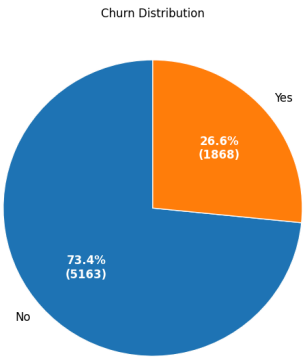
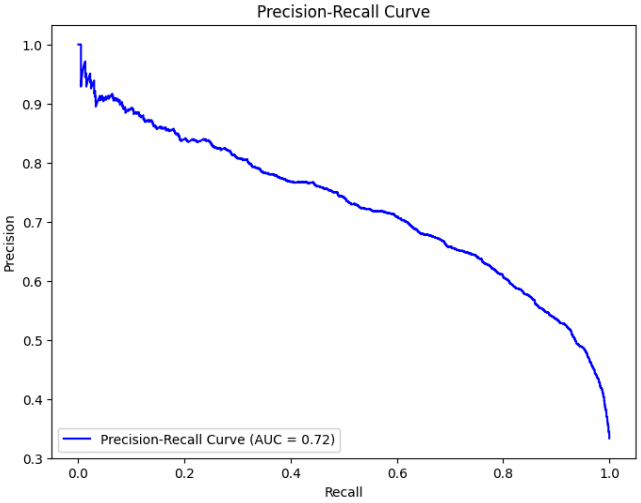
Report

The Telco Customer Churn dataset consists of 21 columns and 7043 rows, each representing a specific customer. Eleven rows with missing values in the TotalCharges column were removed from the analysis.

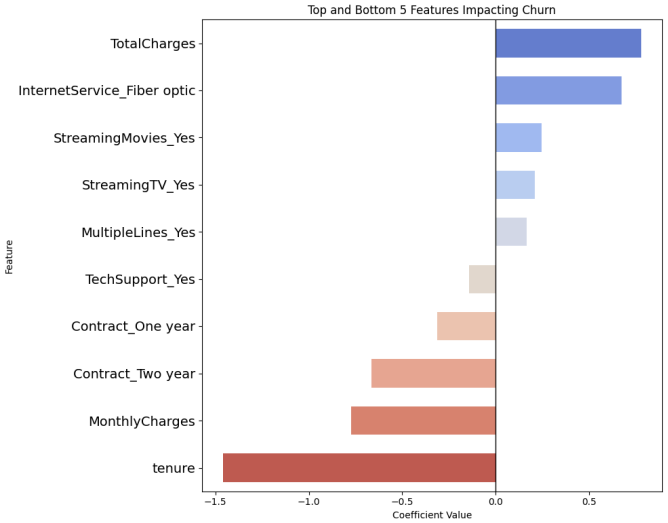


An initial look at the data shows an imbalance in the Churn column, with 5163 non-churners and 1869 churners. Since the data was mostly categorical, numerical values were treated as categorical based on their medians, and Cramér's V was used to evaluate a first relationship between various features and churn. Contract type, tech support, online security, and payment methods stood out as significant factors influencing customer churn. The analysis revealed that customers with month-to-month contracts had a higher tendency to churn compared to those with longer-term contracts. Similarly, the absence of tech support and online security services was associated with increased churn, as was the use of electronic checks compared to other payment methods. Gender had no significant effect on churn rates, with similar patterns observed across male and female customers.

Next, a fine-tuned logistic regression model was built to predict customer churn. The data was transformed using one-hot encoding to handle categorical variables and the imbalance within churn was addressed with SMOTE. Hyperparameters were optimized using grid search, and the model was evaluated using 10-fold cross-validation. The model showed decent balanced accuracy (~77%), though precision (~60%) was lower than recall (~81%), indicating that the model is better at identifying churners than non-churners. This imbalance resulted in a modest F1-score of approximately 69%. The Precision-Recall curve had an AUC of 0.72, further supporting the model's ability to detect churners. Adjustments to the model could be made to focus on improving either precision or recall, but doing so might negatively impact the other.



The logistic regression model also provided insights into the factors influencing churn. Tenure (the number of months the customer has been with the company) was the most critical factor in preventing churn, followed closely by customers with one- or two-years long contracts. The presence of tech support and strong online security services also contributed to customer loyalty. On the other hand, customers with higher total charges were more likely to churn, likely due to unhappiness with pricing and possibly cheaper options from competitors. Customers using fiber optic connections also showed a higher propensity to churn, suggesting issues with this service. Additionally, paperless billing and the use of electronic checks were associated with higher churn rates, indicating that some customers may prefer more traditional payment and billing methods.



In conclusion, the company should focus on customer retention by encouraging customers to subscribe to longer-term contracts. Human interaction also appears to be important, and maintaining a personal touch could help against the skepticism toward automated services, making online security a priority. Finally, improvements to the fiber optic service and continued investment in tech support are essential to reduce churn and enhance customer satisfaction.