Eugene Lacatis, Atharva Prasanna Mokashi, Irwin Salamanca, Matthew Tang *(Group 9)*
Rakesh Ranjan
CMPE 272
28 Sept 2025

Classification Assignment

Background

For the assignment, we are building a classification model to predict customer churn. Two Kaggle datasets were given: Telco Customer Churn Dataset and Customer Personality Analysis. The classification algorithm used to predict customer churn based on the given datasets was a decision tree. Overall, the decision tree classification model performed ok. In both datasets the model had false positives and negatives. The results and key observations for the model on each dataset are presented below.

Results

Telco Customer Churn Dataset

The model was reasonably *accurate (77.83%)*. In terms of *precision*, it was able to predict **58.07%** of customers predicted to churn correctly, but it still had a fair share of false alarms. In terms of *recall*, it was able to identify churned customers at **59.63%**. However, it missed quite a few. The *F1-Score* as a result would be 58.84% as the model struggled with false positives and false negatives. Overall, the *AUC* score was .**8195**, indicating it can effectively differentiate between churners and non-churners.

Some key features contributing to churn prediction were customer tenure, whether the customer subscribed to fiber optic internet service, and total charges (overall spending).

Customer Personality Analysis Dataset

The model was reasonably *accurate (87.16%)*. In terms of *precision*, it was able to predict **72.73%** of customers predicted to churn correctly pretty well. In terms of *recall*, it did not do well in capturing actual churners with a score of **23.88%**. The *F1-Score* as a result would be **35.96%** as the model struggled as there was a poor balance between precision and recall. Overall, the *AUC* score was .**7138**, indicating it can moderately differentiate between churners and non-churners.

Some key features contributing to churn prediction included whether the customer accepted the fifth campaign offer, how recently the customer's site visited, and if the customer accepted the third campaign offer.

Discussion - Improving Customer Retention

Some ways to improve customer retention in regards to the Telco Customer Churn Dataset are to focus on new customers to make sure they have a good first impression and experience, provide discounts, and have competitive internet prices.

Some ways to improve customer retention in regards to the Customer Personality Analysis Dataset are to personalize the campaign offers, improve campaign timing, and to encourage frequent site visits with potentially loyalty programs, exclusive content, or rewards.

Conclusion
Overall, the classification model we performed decently. The decision tree classification algorithm produced both false positives and false negatives with both datasets. The model performed better on the Telco Customer Churn dataset. The model was too conservative on the Customer dataset as it missed too many true positives but avoided false positives whereas the model had a good balance between precision and recall with the Telco dataset.