

Customer Churn Prediction for SaaS CRM

Project Summary:

This project is about predicting churn among customers of a SaaS CRM platform using a synthetic dataset designed to reflect real-world customer behavior. The goal was to identify patterns and build models that can distinguish between customers who are likely to renew their subscription and those who are not.

Dataset Overview:

The dataset includes:

- Customer demographics and subscription metadata (e.g., SignupDate, LastLoginDate, AccountAgeDays)
- Behavioral features like TotalLogins, InvoicesCount, OpportunitiesCount
- Business segment and industry
- Target variable: Churned (1 = churned, 0 = active)

Data Preparation:

- Created synthetic LastLoginDate from LastLoginDaysAgo
- Calculated monthly and total aggregates for user activity
- Cleaned data and removed non-informative columns (e.g., CustomerID)
- Encoded categorical features

EDA (Exploratory Data Analysis):

- Used histograms and scatter plots to examine distributions and outliers
- Correlation heatmaps helped identify redundant features
- Visualized differences in features between churned vs. retained customers

Customer Churn Prediction for SaaS CRM

Models Evaluated:

1. Logistic Regression:

- Simple, interpretable
- Performed the best with high accuracy

2. XGBoost:

- Overfitted quickly on synthetic data
- Didn't generalize well for this dataset

3. Random Forest:

- Good performance and insight into feature importance
- Slightly behind Logistic Regression in predictive power

Model Selection:

Logistic Regression was chosen as the final model because of its simplicity and strong performance.

Even though more complex models were explored, they did not significantly outperform Logistic Regression.

What I Learned:

- Importance of meaningful feature engineering
- Simpler models can outperform complex ones if data is high-quality
- Domain knowledge plays a critical role in shaping predictive models

Next Steps:

- Test models on real-world CRM datasets
- Experiment with imbalance techniques (SMOTE, cost-sensitive learning)
- Consider deployment via Flask or Streamlit

Customer Churn Prediction for SaaS CRM

Author: Pegah Kashani