Video Streaming Churn Project | Feature Engineering (Milestone 4)

Executive Summary Report

Prepared by: PhD Aleksandar Osmanli

Project Overview

I'm currently developing a data analytics project aimed at increasing overall growth by preventing monthly user churn on video streaming service. For the purposes of this project, churn quantifies the number of users who have cancelled the monthly subscription of the video streaming service. Feature engineering offer possibility to enhance the ML models and achieve better results by combining the most predictive basic features into new features, which can add predictive power to the modeling. **This report offers details and key insights from Milestone 4, which impact the future development of the overall project.**

Milestone 4 – Feature Engineering

🎯 **Target Goal:** Create new interactive features and check their additional predictive power on user subscription churn.

🎯 **Methods:**

- Assessment of churn risk among the most predictive basic features
- Creation of interactive features
- Evaluation of their correlation with churn

🎯 **Impact:** By deep engineering, the new features can reveal important relationships and predict the target feature

Correlation of Features with Churn

Feature	Correlation
AccountAgeAverageViewingDuration	-0,2113
AccountAgeViewingHoursPerWeek	-0,2015
AccountAgeContentDownloadsPerMonth	-0,1994
AccountAge	-0,1977
AccountAge_Per_MonthlyCharges	-0,1943
ViewingHoursPerWeekAverageViewingDuration	-0,1728
AverageViewingDurationContentDownloadsPerMonth	-0,1718
TotalChargesAverageViewingDuration	-0,1607
ViewingHoursPerWeekContentDownloadsPerMonth	-0,1605
AverageViewingDuration_Per_MonthlyCharges	-0,1576
TotalChargesViewingHoursPerWeek	-0,1520
TotalChargesContentDownloadsPerMonth	-0,1513
AccountAge_Per_UserRating	-0,1512
AverageViewingDuration	-0,1469
ContentDownloadsPerMonth_Per_MonthlyCharges	-0,1456
ViewingHoursPerWeek_Per_MonthlyCharges	-0,1437
ContentDownloadsPerMonth	-0,1298
ViewingHoursPerWeek	-0,1286
MonthlyCharges_Per_AverageViewingDuration	0,1241
MonthlyChargesSupportTicketsPerMonth	0,1207
TotalCharges	-0,1205
SupportTicketsPerMonth_Per_AverageViewingDuration	0,1189
AccountAge_Per_SupportTicketsPerMonth	-0,1154
AccountAgeWatchlistSize	-0,1152
AverageViewingDuration_Per_UserRating	-0,1145
MonthlyCharges_Per_ViewingHoursPerWeek	0,1104
MonthlyCharges_Per_AccountAge	0,1100
SupportTicketsPerMonth_Per_AccountAge	0,1067
SupportTicketsPerMonth_Per_ViewingHoursPerWeek	0,1060
ContentDownloadsPerMonth_Per_UserRating	-0,1049
ViewingHoursPerWeek_Per_UserRating	-0,1022
TotalCharges_Per_UserRating	-0,1007
MonthlyCharges	0,1005

Next steps

➔ **Taking into account the benefits of tree-based ML models, evaluate different types of models to assess the best roc_auc score on the validation dataset.**