Customer Churn Prediction - Capstone Project Summary

Overview

This capstone project focuses on predicting customer churn in an Iranian telecom company and developing retention strategies. Using a real-world dataset, the study applied logistic regression, K-nearest neighbors, and random forest classifiers to identify key drivers of churn and evaluate model accuracy. Random forest achieved the highest performance, with insights translated into actionable business strategies.

Tools & Techniques

- Programming Language: R

- Libraries: caret, ggplot2, dplyr, randomForest, e1071

- Methods: Data cleaning, normalization, correlation analysis

- Models: Logistic Regression, KNN, Random Forest

- Evaluation: Accuracy, Precision, Recall, F1-score, AUC

Project Highlights

The project helped identify top churn indicators such as service call frequency, data usage, and billing information. Random Forest emerged as the most reliable model, achieving strong predictive performance. Business recommendations included loyalty rewards, proactive service upgrades, and segmentation-based offers.

This project was completed as the final-year capstone for the Master's program in Data Science and Business Analytics at Asia Pacific University (APU), Malaysia.