**Business Requirements Document (BRD)**

**Project Name:** Customer Churn Prediction  
**Prepared by:** James Bell  
**Date:** *4/8/25*

**1. Business Objective**

Retention teams in Customer Success need a data-driven method to predict which active and valued customers are at risk of churning in order to prioritize outreach and improve retention strategies. The goal is to enable proactive engagement that reduces revenue loss and improves customer satisfaction.

**2. Problem Statement**

There is no current predictive mechanism to rank active customers by churn risk. This prevents:

* Timely intervention with high-risk, high-value accounts
* Efficient allocation of retention resources
* Strategic planning to improve overall churn rate
* Understanding why customers are leaving us
* Understand what drives a particular customer to leave us

**3. Goals**

* Build a classification model to predict churn risk using engagement, behavior, and customer attributes
* Use SHAP to identify key drivers of churn across the customer base
* Provide a ranked output of active customers by churn probability and value
* Deliver clear visualizations and written summaries that support stakeholder decisions

**4. Data Sources**

Simulated dataset with 10,000 customer records, containing:

* Features: region, account\_type, device\_type, tenure\_months, monthly\_spend, logins, support\_tickets, and engagement flags
* Target: churned (binary classification)

**5. Deliverables**

* Python notebook with XGBoost modeling pipeline and performance metrics
* SHAP feature importance plots (bar, waterfall, summary)
* Final list of active customers ranked by churn probability × value
* Markdown summary explaining model results and next steps
* Option for individual driver using SHAP for further analysis, should the appetite exist
* GitHub-hosted project with documentation and visual output

**6. Success Criteria**

* Classification model achieves strong ROC AUC and acceptable precision/recall for the churned class
* SHAP results align with business expectations (e.g., free tier churns more, lower engagement drives risk)
* Stakeholders can act on the predictions confidently, starting with top-value at-risk customers

**7. Assumptions**

* Only known data at scoring time is used for predictions (no leakage from future)
* Model is designed for monthly churn prediction, but can be adapted for quarterly
* CLV (value) is approximated using monthly spend × tenure or other derived metrics