Predictive Analytics for Customer Retention in E-Commerce and Banking  
  
EMINE UYSAL (ID: 501304049)

Toronto Metropolitan University

CIND 820: Big Data Analytics Project

May 19th, 2025

**ABSTRACT**

Customer churn comprises a persistent challenge in data-driven industries such as e-commerce and banking, where customer retention significantly influences revenue stability and long-term growth. This project investigates churn dynamics by applying predictive analytics to identify behavioral patterns related to customer churn and to develop targeted retention strategies.

The analysis is based on two open-source datasets obtained from Kaggle: the “*Customer Behavior in E-commerce”* dataset and the “*Bank Customer Churn*” dataset. The first includes online retail behavior such as visit frequency, average order value, and transaction history, while the latter provides demographic and account data from banking customers. These datasets allow for sector-specific insights and a comparative analysis of churn predictors across domains.

Based on my prior professional experience in e-commerce analytics, this study applies machine learning models — Logistic Regression, Random Forest, and XGBoost — to predict churn based on historical behavior. The methodology includes data preprocessing, feature engineering, and customer segmentation. Model performance will be evaluated using classification metrics including accuracy, precision, recall, and F1-score.

To complement the predictive models, an interactive Power BI dashboard will visualize key findings, enabling users to explore churn patterns, identify high-risk customer segments, and compare industry-specific drivers. The dashboard is designed to support both technical and non-technical stakeholders in making informed, data-driven decisions.

The expected outcome of the project is a set of actionable recommendations for improving customer retention strategies in each sector, along with a framework for implementing scalable churn prediction systems. Ultimately, the project demonstrates how machine learning and business intelligence tools can be used to address real-world challenges in customer lifecycle management.

*Note: Both datasets provide useful behavioral and demographic information but have limited scope and synthetic elements, which may affect generalizability. Future work aims to include larger, real-world datasets to enhance model robustness and relevance.*