

Customer Churn Prediction in Bank of Kigali

BSc. in Software Engineering

Student Name: Steven Shyaka

Supervisor: Pelin Mutanguha

Institution: African Leadership University - Software Engineering

Date: October 10, 2025

Abstract

Customer retention is a strategic challenge for commercial banks, particularly as Rwanda's financial sector rapidly digitizes. The Bank of Kigali (BK) experiences customer churn due to changing expectations and digital competition. This project develops a machine learning (ML)-based churn prediction system tailored for BK using a synthetic yet realistic dataset of 150,000 anonymized records. Logistic Regression, Random Forest, and XGBoost models will be trained, evaluated using Accuracy, Precision, Recall, F1-Score, and AUC-ROC metrics. SHAP (SHapley Additive exPlanations) will provide interpretable insights into key churn drivers. The system will integrate with dashboards for Retention Officers and Analysts to proactively identify at-risk customers and guide retention campaigns, promoting data-driven decision-making in Rwandan banking.

Table of Contents

- List of Tables
 - List of Figures
 - List of Acronyms/Abbreviations
 - Chapter One: Introduction
 - Chapter Two: Literature Review
 - Chapter Three: System Analysis and Design
 - Chapter Four: Implementation
 - Chapter Five: Results and Discussion
-

List of Tables

- Table 1: Summary of Data Features
 - Table 2: Model Performance Comparison
 - Table 3: Research Budget
 - Table 4: Research Timeline
 - Table 5: User Role Responsibility Matrix
-

List of Figures

- Fig 1: ML Pipeline Diagram
 - Fig 2: Use Case Diagram – Retention Officer
 - Fig 3: Use Case Diagram – Retention Analyst
 - Fig 4: Use Case Diagram – Admin
 - Fig 5: Activity Diagram – Individual Prediction
 - Fig 6: Activity Diagram – Bulk Prediction & Campaign Trigger
 - Fig 7: Sequence Diagram – Customer Churn Prediction
 - Fig 8: Class Diagram – Customer & Prediction Model
 - Fig 9: Master Workflow Activity Diagram
-

List of Acronyms/Abbreviations

Acronym	Meaning
ML	Machine Learning

EDA Exploratory Data Analysis

AUC-ROC Area Under Curve – Receiver Operating Characteristics

SHAP SHapley Additive exPlanations

TBD To be Determined

Chapter One: Introduction

1.1 Background

Customer retention is a strategic challenge for commercial banks, especially as Rwanda's financial sector rapidly digitizes. Bank of Kigali (BK) faces churn due to changing customer expectations and digital competition. While loyalty programs exist, predictive analytics enables proactive retention. This project uses ML to develop a locally tailored churn prediction system for BK.

1.2 Problem Statement

Existing global tools (SAS, IBM SPSS) are costly, complex, and not localized. Rwanda lacks empirical research on ML-based churn prediction in banking. BK's data allows development of an interpretable, actionable model to predict churn and guide retention interventions.

1.3 Objectives

General Objective:

Develop a ML-based churn prediction system for Bank of Kigali.

Specific Objectives:

- Analyze BK customer data to identify churn drivers.
- Build ML models: Logistic Regression, Random Forest, XGBoost.
- Evaluate model performance: Accuracy, Precision, Recall, F1, AUC-ROC.
- Provide interpretable insights using feature importance and SHAP.
- Integrate system functionalities for Retention Officers, Analysts, and Admins.

1.4 Research Questions

- Which features most accurately predict customer churn at BK?
- Can ML models provide actionable insights for retention campaigns?
- How should the system differentiate single-customer vs. bulk prediction use cases?

1.5 Project Scope

- Focus on BK only.
- Use anonymized synthetic data (150,000 records) reflecting BK operations.
- Implement ML models, dashboards, and prediction interfaces.
- Include UI for Retention Officers (individual predictions) and Analysts (bulk campaigns).

1.6 Significance

- Proactively identify at-risk customers.
- Guide retention campaigns and promotions.
- Facilitate adoption of ML in Rwandan banking.

1.7 Dataset and Features

Notes on Dataset:

The dataset was synthetically generated using BK's structure, schema, and data distribution patterns. No personally identifiable information (PII) was used, ensuring compliance with data protection regulations. Data sources are based on BK operational data characteristics from 2019–2025.

Feature	Description	Type
Customer_ID	Unique customer identifier	Numeric
Account_Number	Unique bank account number	Numeric
Gender	Customer gender (Male/Female)	Categorical
Age	Customer age in years	Numeric

Nationality	Customer nationality (mostly Rwandan, some foreign)	Categorical
Account_Type	Type of account (Savings, Current, Fixed Deposit, etc.)	Categorical
Branch	Branch where the account is held	Categorical
Currency	Account currency (RWF, USD, EUR, CAD, CHF)	Categorical
Balance	Current or average account balance	Numeric
Tenure_Months	Number of months since account opening	Numeric
Num_Products	Number of BK products held (loans, cards, mobile, etc.)	Numeric
Has_Credit_Card	Whether customer owns a BK credit card (1 = Yes, 0 = No)	Binary
Account_Status	Account activity status (Active, Inactive, Dormant, Dom. Closed, Unclaimed)	Categorical
Last_Transaction_Date	Date of most recent transaction	Date
Account_Open_Date	Date account was opened	Date
Transaction_Frequency	Average number of monthly transactions	Numeric
Average_Transaction_Value	Mean value per transaction	Numeric
Mobile_Banking_Usage	Frequency of mobile app or USSD usage per month	Numeric
Branch_Visits	Average physical branch visits per year	Numeric
Complaint_History	Number of complaints or service issues reported	Numeric

Churn_Flag	1 = churned (Dom. Closed / Unclaimed), 0 = retained	Binary
------------	---	--------

Chapter Two: Literature Review

2.1 Local Studies

- Hagenimana & Kengere (2025): Online banking improves satisfaction.
- Nyiranzabamwita & Harelimana (2019): E-banking improves service delivery.
- Mutesi (2022): Electronic banking positively affects customer retention.
- Yoronimu & Tumutegye (2018): CRM practices influence profitability.

2.2 International Studies

- Asfaw (2023): ML predicts churn in Ethiopian banks.
- Molla et al. (2025): ML outperforms traditional methods in Awash Bank.
- Basri (2025): Random Forest vs Logistic Regression for churn.
- Kabbar & Herath (2025): Multi-algorithm ML improves retention strategies.

2.3 Research Gap

Limited research in Rwanda applies ML to actual churn prediction, leaving a gap for actionable, local models using real bank data.

Chapter Three: System Analysis and Design

3.1 Actors

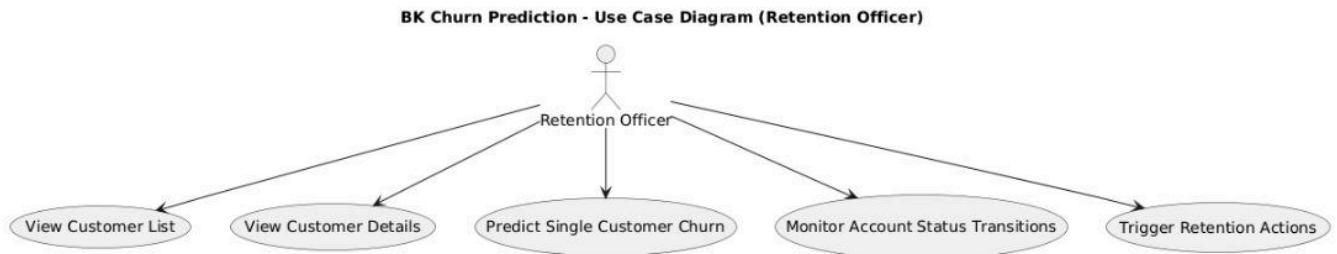
- **Retention Officer:** Inputs single Customer ID, receives churn prediction and risk status, views detailed customer history, triggers retention recommendations.
- **Retention Analyst:** Executes bulk predictions, segments customers by risk levels and product usage, triggers retention campaigns.

- **Admin:** Manages system users, approves campaigns, monitors performance.
- **IT/Data Engineer:** Supports retraining, deployment, and data updates.

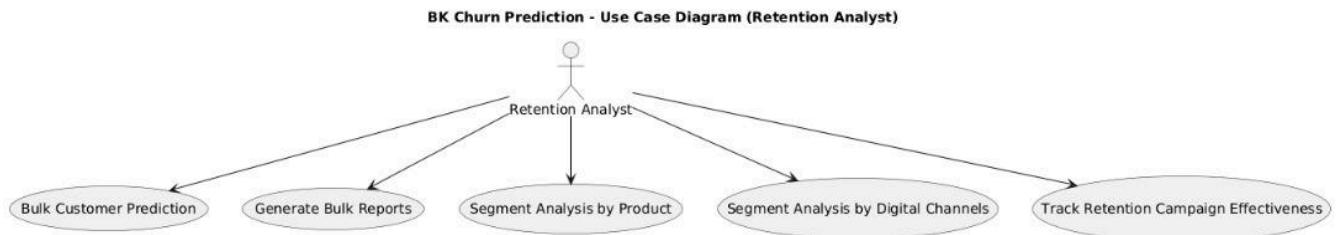
3.2 Use Cases

Placeholders for diagrams:

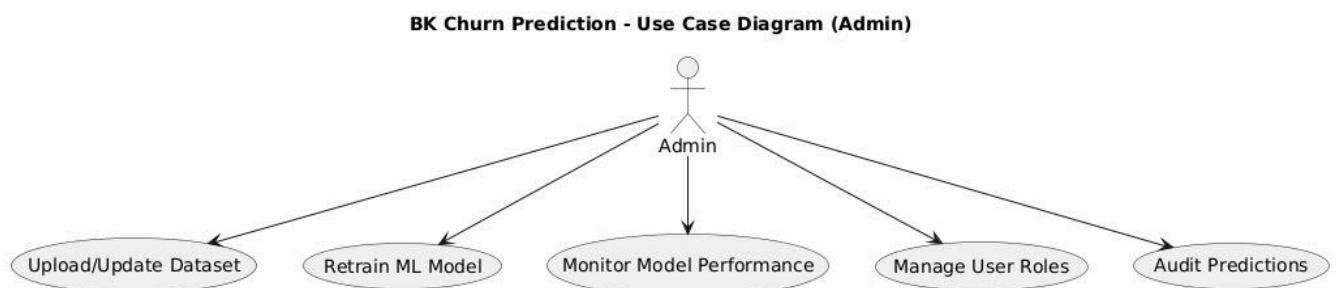
- Fig. 2: Use Case – Retention Officer



- Fig. 3: Use Case – Retention Analyst



- Fig. 4: Use Case – Admin



Example (Retention Officer):

- Menu → Customers → Table view of all customers
- Click → View Details → Shows full info, products, churn prediction, and projected status changes

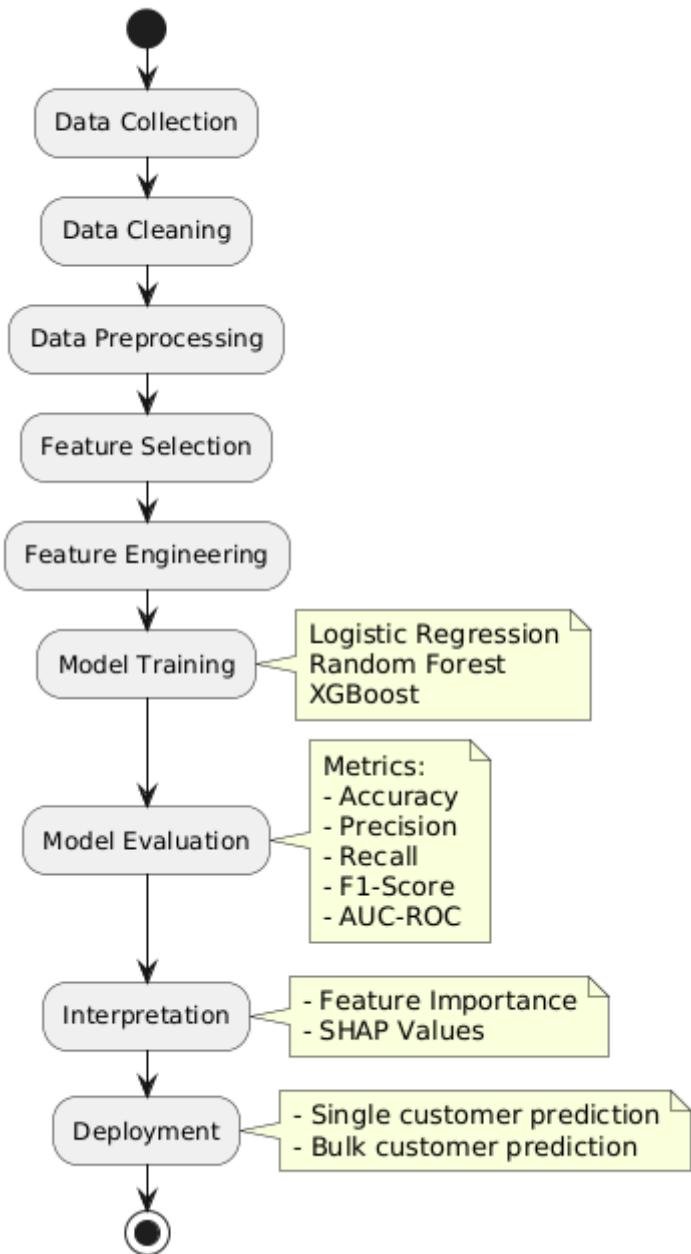
Customer Status Definitions:

Status Transition	Description
Active → Inactive	No transactions in 6 months
Inactive → Dormant	6–12 months without transactions
Dormant → Dom. Closed	12 months–4 years
Dom. Closed → Unclaimed	4+ years without activity

3.3 ML Model Workflow

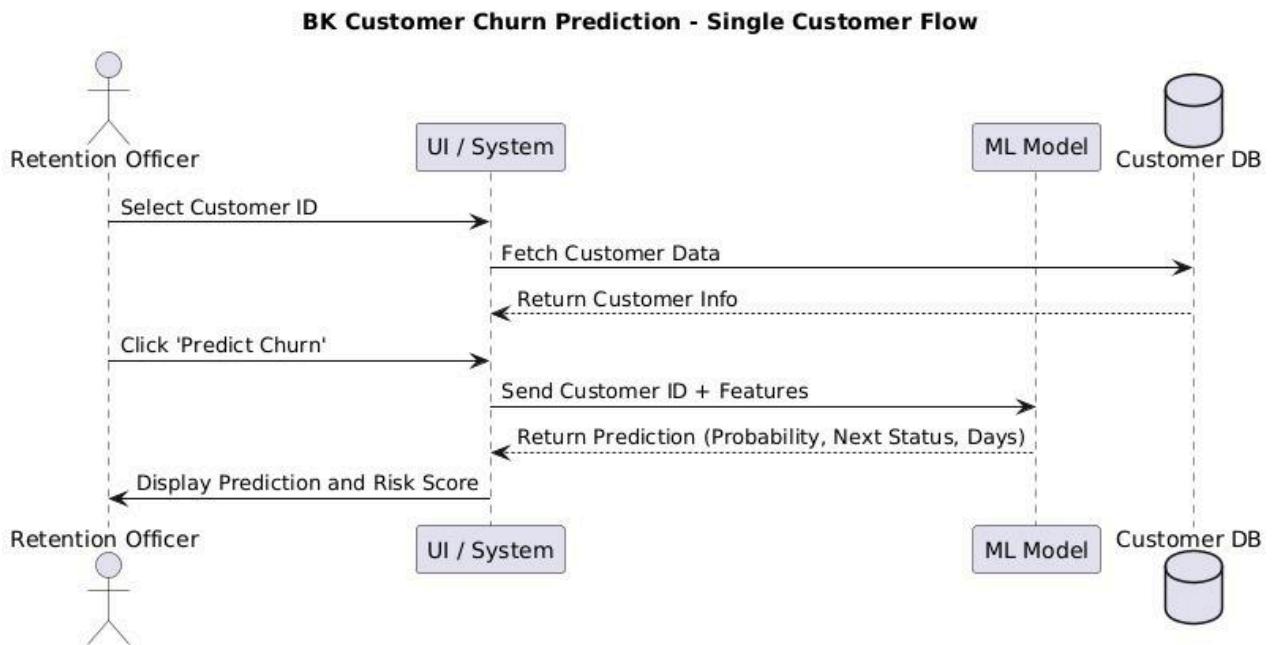
Fig 1: ML Pipeline - Customer Churn Prediction

ML Pipeline - Customer Churn Prediction

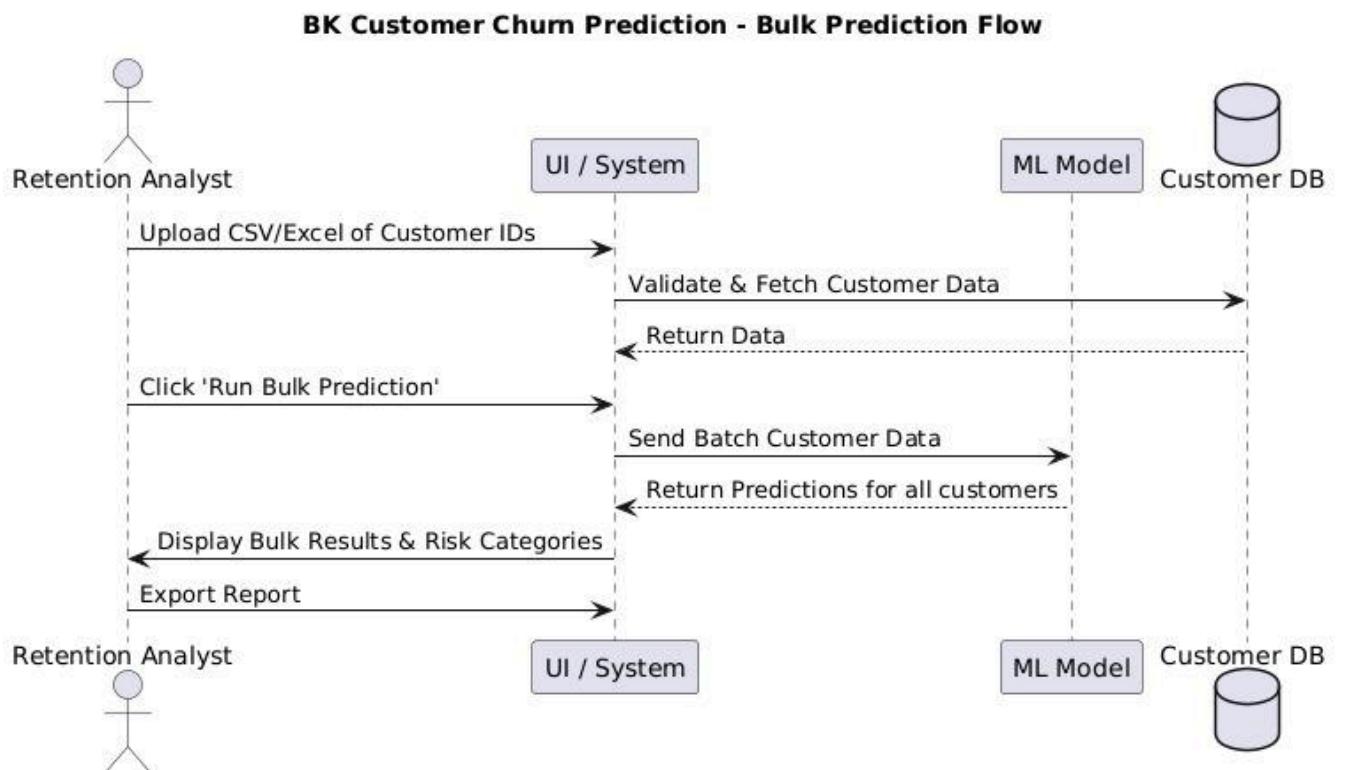


3.4 Activity Diagrams

- Fig. 5: Individual Customer Prediction (Retention Officer)

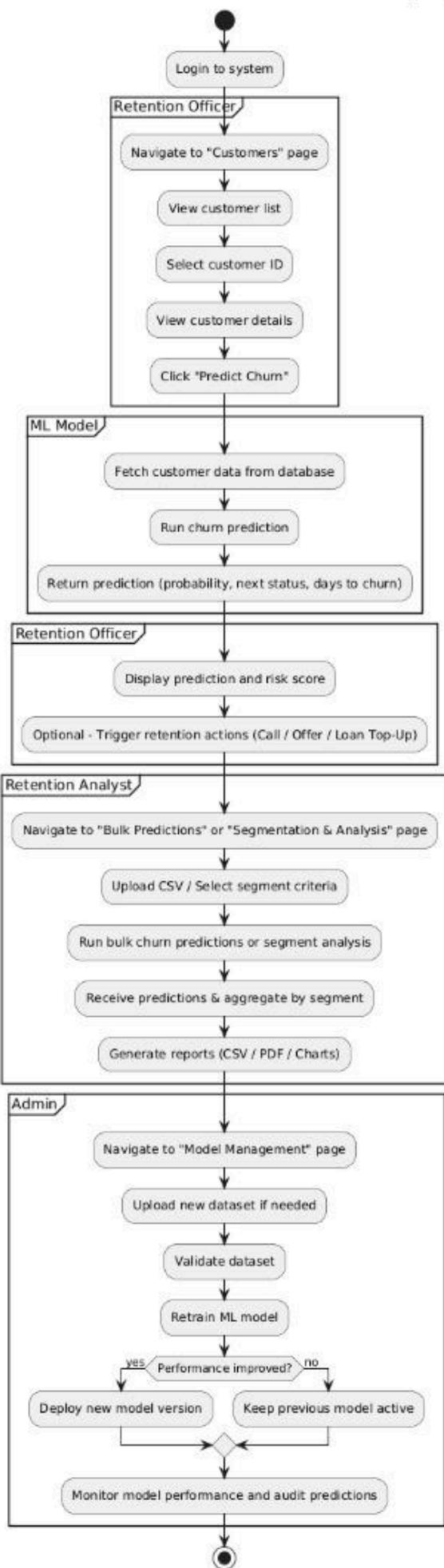


- Fig. 6: Bulk Prediction & Campaign Trigger (Retention Analyst)

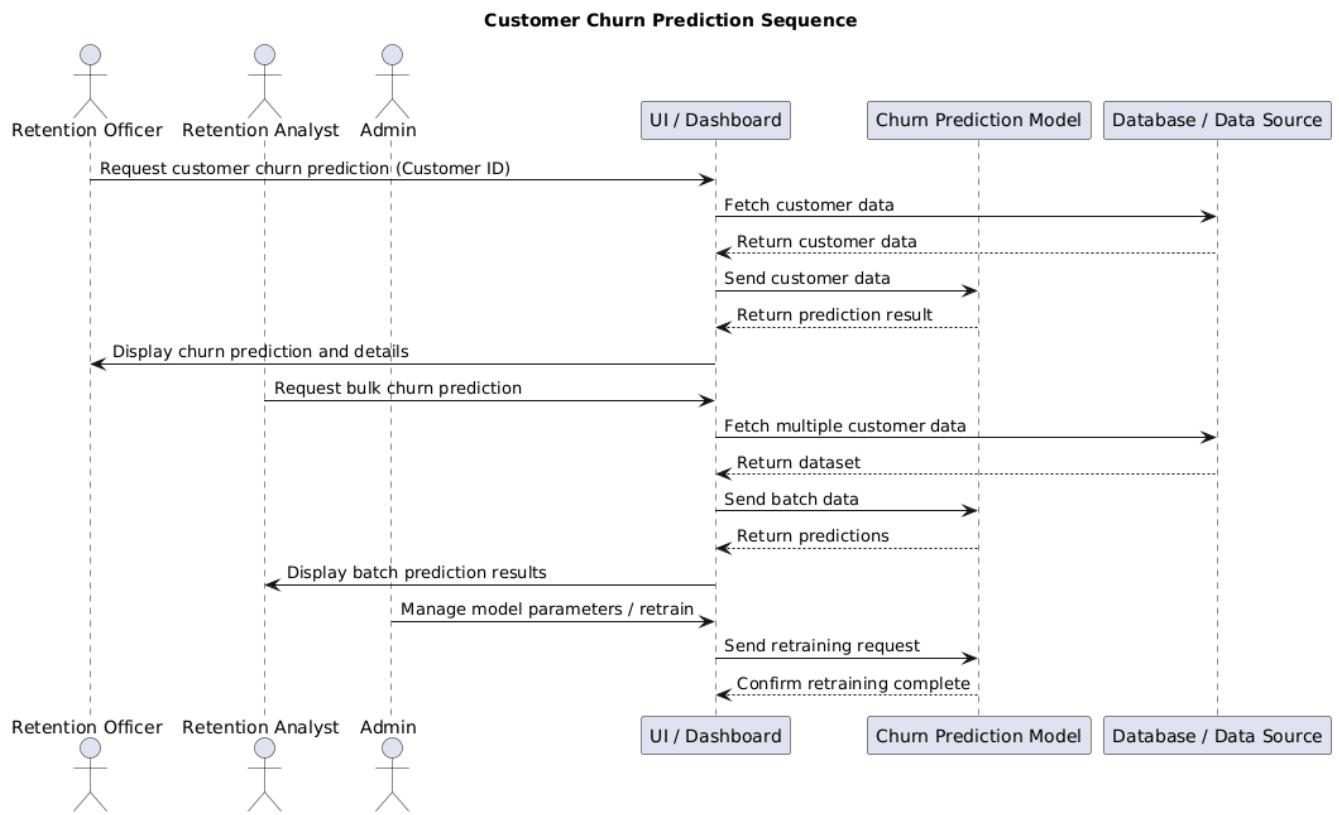


- Fig. 9: Master Workflow (All actors)

BK Customer Churn Prediction - Master Workflow Activity Diagram

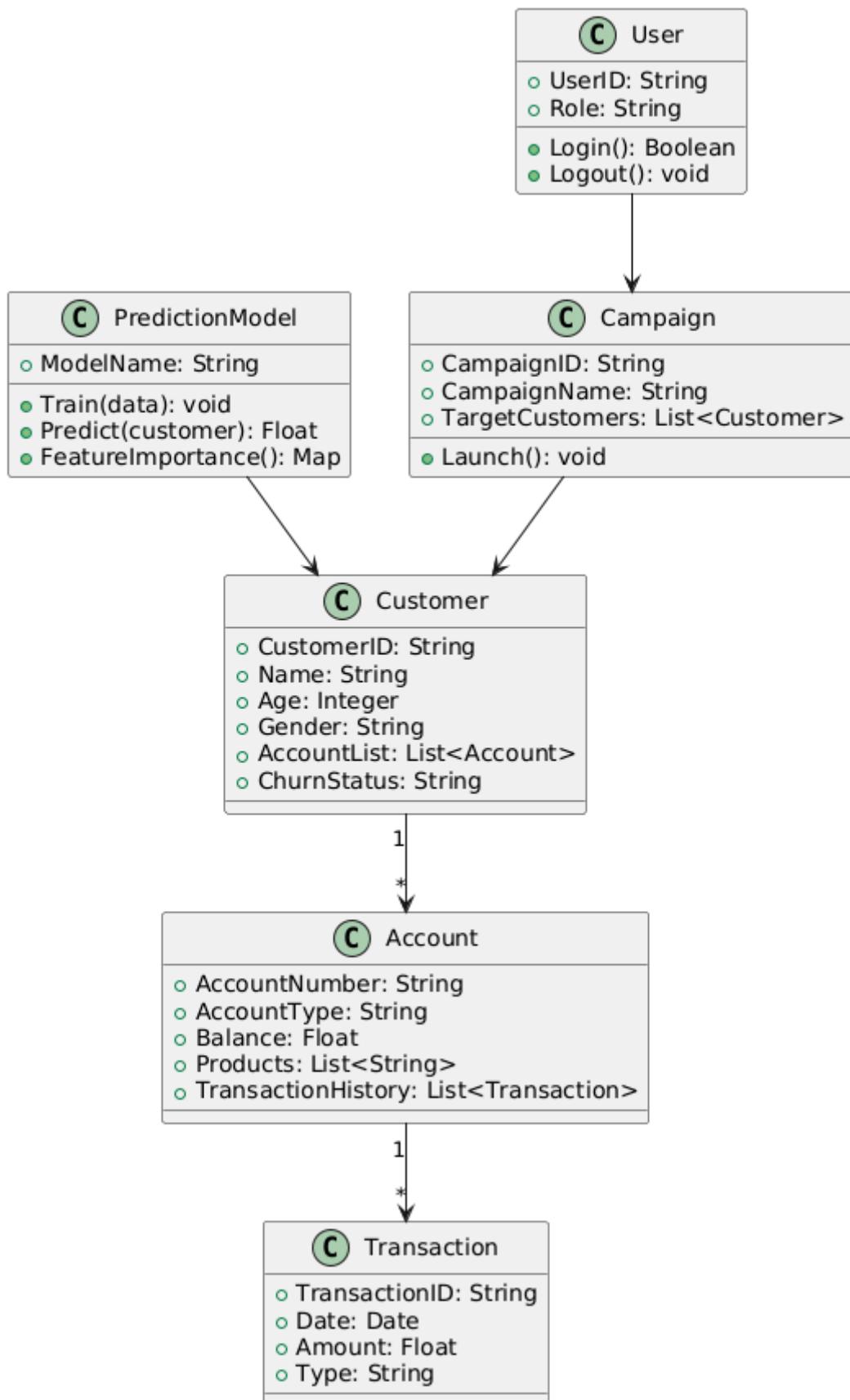


3.5 Sequence Diagram – Customer Churn Prediction



3.6 Class Diagram – Customer Churn System

Customer Churn Prediction System - Class Diagram



3.7 Tools & Technologies

- **Language:** Python
 - **Libraries:** scikit-learn, XGBoost, TensorFlow, Keras, SHAP
 - **Visualization:** Matplotlib, Seaborn
 - **Storage:** PostgreSQL / CSV
 - **Version Control:** GitHub
-

Tables

Table 2: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	TBD	TBD	TBD	TBD	TBD
Random Forest	TBD	TBD	TBD	TBD	TBD
XGBoost	TBD	TBD	TBD	TBD	TBD

Table 3: Research Budget

Item	Description	Estimated Cost (RWF)
Data Collection	Data acquisition & cleaning	50,000
Software Tools	Python libraries & hosting	30,000
Hardware	Computers & storage	60,000
Miscellaneous	Printing, transport, meetings	20,000
Total		160,000

Table 4: Research Timeline

Week	Activity
Week 1–2 (01–12/09/2025)	Write proposal
Week 3 (15–19/09/2025)	Proposal presentation & revisions
Week 4–5 (22/09–03/10/2025)	Data preparation & EDA
Week 6–7 (06–17/10/2025)	Model training & tuning
Week 8 (20–24/10/2025)	Model evaluation & feature interpretation
Week 9 (27–31/10/2025)	Draft Chapter 4: Implementation
Week 10–11 (03–14/11/2025)	Draft Chapter 5: Results & Discussion
Week 12 (17–21/11/2025)	Prepare final report
Week 13 (24–28/11/2025)	Final defense & submission

Table 5: User Role Responsibility Matrix

Function / System Feature	Retention Officer	Retention Analyst	HOD / Manager	Admin
View customer details	✓	✓	✓	✓
Run single-customer prediction	✓	✓	✓	✓
View churn probability & explanation (SHAP insights)	✓	✓	✓	✓
Update customer retention notes / feedback	✓	✓	✓	✓
Bulk churn prediction (segment-level)	✗	✓	✓	✓
Trigger retention campaigns	✗	✓	✓	✓
View campaign performance / conversion rates	✓	✓	✓	✓
Adjust churn thresholds (risk categorization)	✗	✗	✓	✓
Approve or review retention campaigns	✗	✓	✓	✓
View model performance metrics (accuracy, AUC, etc.)	✗	✓	✓	✓
Monitor customer segment trends (active/inactive/dormant)	✓	✓	✓	✓
Retrain the model with new data	✗	✗	✗	✓
Deploy updated model to production	✗	✗	✗	✓
Manage user access and permissions	✗	✗	✗	✓
Audit and compliance reporting	✗	✗	✓	✓

References

- 1. Hagenimana, T., & Kengere, T. (2025).**
Online Banking Services and Customer Satisfaction in Rwanda: A Case Study of the Bank of Kigali.
International Journal of Applied Business and Social Science (IJABS), 1(1), 1–15.
Available at: <https://www.besra-journals.net/index.php/ijabs/article/view/11>
- 2. Nyiranzabamwita, J., & Harelimana, J. B. (2019).**
The Effect of Electronic Banking on Customer Services Delivery.
Enterprise Risk Management (ERM), 5(1), 15904.
Available at: <https://www.macrothink.org/journal/index.php/erm/article/view/15904/0>
- 3. Mutesi, J. C. (2022).**
Effect of Electronic Banking on Customer Satisfaction in Rwanda: Case of Bank of Kigali Headquarter.
Scholars Journal of Economics, Business and Management (SJEBM), 9(12), 452–460.
Available at: <https://www.saspublishers.com/article/4624/>
- 4. Yoronimu, C., & Tumutegye, R. (2018).**
CRM and Bank Profitability: BK Nyagatare Branch.
KIU Repository.
Available at: <https://ir.kiu.ac.ug/items/913f1243-121f-4cce-95c5-e05b08c10aa2>
- 5. Asfaw, T. (2023).**
Customer Churn Prediction in Ethiopian Banks Using Machine Learning.
The Scientific Temper, 14(3), 618–624.
Available at: <https://scientifictemper.com/index.php/tst/article/view/678>
- 6. Molla, A. M., Yimer, M. A., & Woldehana, Y. D. (2025).**
Customer Churn Prediction Using Machine Learning Techniques: Awash Bank Wolaita Sodo Region.
Journal of Emerging Computer Technologies (JECT), 5(1), 36–46.
Available at: <https://dergipark.org.tr/en/pub/ject/issue/87666/1623937>
- 7. Basri, M. (2025).**
Logistic Regression vs Random Forest for Banking Churn.
Ultimatics, 17(1), 72–81.
Available at: <https://ibimapublishing.com/articles/JSSD/2025/786386/>
- 8. Kabbar, E., & Herath, N. (2025).**
Seven ML Algorithms for Banking Churn Prediction.
Journal of Software Systems Development (JSSD), Article ID 786386.
Available at: <https://ibimapublishing.com/articles/JSSD/2025/786386/>