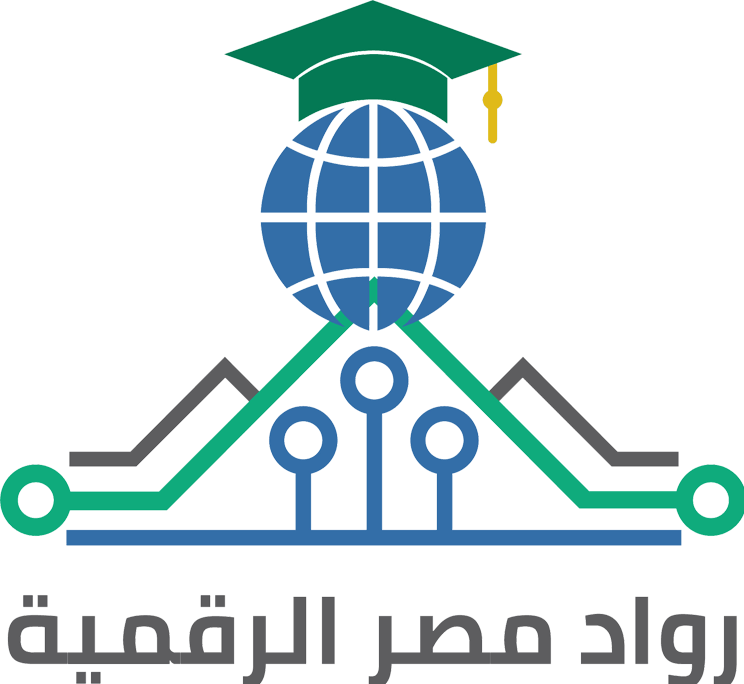
****

**Customer Churn Prediction & Analysis Final Report**

**Done by**: Mariam Ahmed

Noureen Tamer

Haneen Elsayed

Hager Ibrahim

Hazem Metwalli

Youssef Ayman

## Project Summary

### Problem Definition

Customer churn -when a customer discontinues their relationship with a business is a significant issue in customer lifecycle management. The goal of this project is to build a machine learning model that predicts customer churn risk using behavioral and demographic features. Early prediction enables businesses to retain valuable customers and reduce revenue loss.

### Data Exploration

The dataset consists of 23 columns capturing a wide variety of customer information:

- **Demographics:** age, gender, region\_category

- **Account & Behavior:** membership\_category, preferred\_offer\_types, medium\_of\_operation, internet\_option , avg\_transaction\_value, avg\_time\_spent, avg\_frequency\_login\_days, points\_in\_wallet

- **Engagement & Feedback**: used\_special\_discount, past\_complaint, feedback, complaint\_status

- **Target Variable:** churn\_risk\_score (binary: 0 = low churn risk, 1 = high churn risk)

### Key Insights from EDA

**1.** **Wallet Balance Matters Most**

The feature **(points\_in\_wallet)** is the most influential predictor of churn, suggesting that customers with higher wallet balances are significantly less likely to churn.

**2. Membership Category Strongly Predicts Churn**

- Customers with **No Membership** have the highest churn risk (very high positive T-statistic).

- Premium, Platinum, Gold, and Silver memberships are associated with lower churn rates.

**3. Customer Feedback is Highly Predictive**

Several feedback types are statistically significant in explaining churn behavior:

- Negative feedback such as **Poor Website, Poor Product Quality, and Poor Customer Service** are strongly correlated with high churn.

- **Positive feedback** like User Friendly Website, Reasonable Price, and Quality Customer Care are linked with customer retention.

**4. Login Behavior and Engagement**

- **avg\_frequency\_login\_days** and **avg\_time\_spent** are significant indicators. Customers who log in more frequently and spend more time tend to stay.

- **engagment\_score** is also significant — more engaged customers are less likely to churn.

**5. Transaction Behavior**

**- avg\_transaction\_value** is significant: customers with higher spending are less likely to churn.

**6. Referral & Offers Influence Churn**

- Customers not referred by others are more likely to churn.

- Preferred offers such as Gift Vouchers/Coupons reduce churn risk, while customers preferring No offers are more likely to leave.

**7. Device and Connectivity Channel**

- Users operating via Smartphones and those with Wi-Fi connections show slightly higher retention than others.

**8. Recency Metrics**

- “recency\_of\_last\_activity” and “days\_since\_last\_login” are statistically significant: long inactivity periods are linked with higher churn risk.

### Preprocessing Steps

- Missing values imputed

- Categorical features encoded using LabelEncoder or One-Hot Encoding

- Numerical features scaled using a standard scaler

## 2. Model Development

Multiple models were evaluated:

- Logistic Regression

- Random Forest

- XGBoost (final choice)

### Champion Model: XGBoost

This model delivered the best performance on the test data:

- **Accuracy:** 95.09%

- **Precision:** 95.19%

- **Recall:** 94.94%

- **F1 Score:**95.04%

- **AUC:** 98.50%

### Model Artifacts:

- Serialized using pickle

- Files: “best\_xgb\_model.pkl”, “scaler.pkl”

### Feature Importance:

Important features included:

- points\_in\_wallet

- membership\_category

- feedback

- avg\_transaction\_value

- avg\_frequency\_login\_days

- engagment\_score

## 3. Deployment Strategy

The project adopts a dual-deployment architecture for accessibility and scalability:

### A. Streamlit Application

- **Purpose:** Business-friendly UI for churn prediction

**- Features:**

- Real-time prediction on input data

- Simple interface for non-technical users

- **Log File:”**streamlit\_app.log”

### B. FastAPI Application

**- Purpose:** Production-grade REST API

**- Features:**

- predict endpoint with input validation

- Swagger documentation at “docs”

- Web UI for JSON-based predictions

- **Log File:**”fastapi\_app.log”

### Containerization

- Dockerized FastAPI app using a “Dockerfile”

- Deployed on port **9000**

- Entry: “uvicorn fastapi\_app:app --reload --port 9000”

## 4. Business Implications

The ability to predict churn offers tangible business value:

- **Customer Retention:** Enables proactive intervention strategies

**- Resource Allocation:** Focus retention efforts on high-risk segments

**- Revenue Impact:** Reduces cost of customer loss and boosts loyalty

- **Product Feedback Loop:** Uses churn patterns to improve services

## 5. Key Insights, Challenges, and Strategic Decisions

### Insights

- Behavioral traits (wallet balance, feedback, login frequency) are more predictive than demographic factors.

- Membership tier and preferred offers are strong churn signals.

- Negative feedback categories are consistently tied to high churn risk.

### Challenges

- Handling missing values in “points\_in\_wallet” and “referral\_id”

- Dropping uninformative or high-missing-rate features (e.g., “referral\_id”)

- Ensuring consistent encoding and preprocessing during deployment

### Decisions

- Selected XGBoost for its high performance and interpretability

- Designed a dual deployment strategy for business and technical users

- Applied Docker to ensure consistency in deployment environments

## 6. Future Enhancements

- **Model Monitoring** : Add drift detection and performance tracking in production

- **Versioning:** Use MLflow or DVC for model tracking

- **Auto-Retraining Pipelines:** Incorporate CI/CD for retraining with fresh data

**- Cloud Deployment:** Scale via AWS/GCP with load balancing

**- A/B Testing:** Compare multiple model versions in production