# Task 4: GlobalMart - Advanced Churn Prediction

## Internship at Nexus AI Digital — Data Science & Machine Learning
## Author: Syed Huzaifa Bin Khamis

## Overview:

This project predicts **customer churn** (whether a customer will leave or stay) using the **Telco Customer Churn dataset**.

Churn prediction is crucial for businesses because retaining customers is more cost-effective than acquiring new ones.

## Tasks Completed:

## Task 3: Baseline Churn Prediction

- Built a **Logistic Regression model** as a baseline.

- Performance: 80% accuracy, but **weak recall for churners** (missed many who actually churned).

## Task 4: Advanced Churn Prediction

- Improved the model using:

  **- Random Forest**

  **- XGBoost (Extreme Gradient Boosting)**

- Applied feature scaling, label encoding, and handled missing values.

- Evaluated models using **Precision, Recall, F1-Score, and Confusion Matrix.**

## Model Comparison:

| Model | Accuracy | Precision | Recall | F1-Score |

|---|---|---|---|---|

| Logistic Regression | ~80% | Moderate | Low | Moderate |

| Random Forest | ~82% | Better | Improved | Better |

| XGBoost | ~85%+ | High | Best | Best |

**XGBoost outperformed** both **Logistic Regression and Random Forest**, especially in Recall (catching churners), which is critical for business.

## Technologies Used:

- Python

- Pandas, NumPy

- Scikit-learn

- XGBoost

- Matplotlib, Seaborn

## Recommendation:

- Use **XGBoost** for deployment because:

  - It balances accuracy and recall.

  - It reduces false negatives (fewer churners are missed).

  - It's scalable for large datasets.

## How to Run:

1. Install dependencies:

   ```bash

   pip install pandas numpy matplotlib seaborn scikit-learn xgboost.
   ```