E-Commerce Customer Churn Prediction

Name: Arepalli Venkata Lakshmi

Course: Data Analytics / Machine Learning

Business Problem

- E-commerce companies lose revenue due to customer churn.

- Acquiring new customers is 5–7x more expensive than retaining existing ones.

- Goal: Predict which customers are likely to churn in the next 3 months.

- This enables proactive retention strategies and cost savings.

Stakeholders & Business Impact

- Marketing team – targeted retention campaigns

- Business managers – revenue protection & ROI improvement

- Data team – automated churn monitoring

- Outcome: Reduced churn and improved customer lifetime value

Dataset Overview

- Dataset: Online Retail Transactions
- Period: Dec 2009 – Dec 2011
- Size: ~540K transaction records
- Key fields: CustomerID, InvoiceDate, Quantity, UnitPrice, Country
- Challenges: Missing IDs, cancellations, returns, outliers

Data Cleaning & Preprocessing

- Removed missing CustomerID records
- Excluded cancelled invoices and negative quantities
- Handled zero/negative prices
- Outlier removal using IQR method
- Final retention rate ~60%

Feature Engineering

- Converted transaction data → customer-level features
- RFM Features: Recency, Frequency, Monetary value
- Behavioral: Basket size, purchase intervals
- Temporal: Lifetime, recent activity (30/60/90 days)
- Target: Churn defined using 3-month observation window

Models Evaluated

- Logistic Regression (Baseline)

- Decision Tree

- Random Forest

- XGBoost (Gradient Boosting)

- Neural Network (MLPClassifier)

Model Performance Comparison

- Evaluation Metrics: ROC-AUC, Precision, Recall, F1-Score

- Best Model: Random Forest

- ROC-AUC ≈ 0.74

- Balanced precision and recall for churn detection

Business Impact

- Churn rate reduced by targeted retention
- Retention cost: £10 per customer
- Average customer value: £500
- Model enables high ROI by focusing on true churners

Deployment – Streamlit App

- Model deployed using Streamlit Community Cloud
- Features: Single & batch prediction
- Interactive dashboard for stakeholders
- Live URL:
- https://ecommerce-churn-prediction-4mqebfbjtxfjcmampcblqb.streamlit.app/

Key Learnings

- End-to-end ML pipeline design

- Handling real-world messy data

- Importance of churn definition

- Model comparison and evaluation

- Deploying ML models as web apps

Future Improvements

- Hyperparameter tuning for better performance

- Advanced customer segmentation

- Real-time data integration

- Automated retraining and monitoring

Thank You