E-Commerce Customer Churn Prediction

A.N.D.Sripriya

STUDENT ID:23A91A4403

BUSINESS PROBLEM

- Customer churn is one of the most critical challenges faced by e-commerce businesses. When customers stop purchasing, businesses lose recurring revenue and long-term value. Acquiring new customers is significantly more expensive than retaining existing ones. Therefore, identifying customers who are likely to churn in advance is essential. This project focuses on predicting customer churn using historical transaction data to enable proactive retention strategies

STAKEHOLDERS AND BUSINESS IMPACT

- The solution is designed to support multiple business stakeholders by providing actionable insights into customer behavior. Marketing teams can design targeted retention campaigns for high-risk customers Business managers can make data-driven strategic decisions Retention teams can engage customers before churn occurs Overall reduction in customer loss and improved customer lifetime value

DATASET OVERVIEW

- The project uses the UCI Online Retail Dataset, which represents real-world transactional data from an online retail business. The dataset includes detailed purchase information such as invoice numbers, product details, quantities, prices, and timestamps. Total records: 541,909 transactions Time period: December 2010 to December 2011 Source: Kaggle (UCI Online Retail Dataset)

DATA CLEANING CHALLENGES

- Raw transactional data often contains inconsistencies and errors that can negatively affect model performance if not handled properly. Several data quality issues were identified and addressed during preprocessing. Missing Customer IDs and incomplete records Cancelled invoices that do not represent real purchases Negative quantities and invalid prices Duplicate and noisy transactions

FEATURE ENGINEERING

- Transaction-level data was aggregated to customer-level features to better represent individual customer behavior. Feature engineering focused on capturing purchasing patterns and customer engagement over time. Recency: number of days since last purchase Frequency: total number of purchases Monetary: total amount spent Behavioral and temporal features reflecting customer activity

MODELS EVALUATED

- Multiple machine learning models were trained and evaluated to identify the most suitable approach for churn prediction. Each model was compared using consistent evaluation metrics. Logistic Regression ,Decision Tree ,Random Forest ,Gradient Boosting

FINAL MODEL PERFORMANCE

- Gradient Boosting was selected as the final model due to its strong predictive performance and stability. The model demonstrated a good balance between precision and recall, making it suitable for business use.
- ROC-AUC: 0.75
- Precision: 0.74
- Recall: 0.69
- F1-Score: 0.71

BUSINESS IMPACT AND RECOMMENDATIONS

- The churn prediction model enables businesses to take timely and informed actions to retain customers. Predicted churn probabilities can be directly used for marketing and engagement strategies.
- Early identification of high-risk customers
- Personalized offers and loyalty programs
- Improved customer lifetime value
- Better return on marketing investment

DEPLOYMENT

- The final model was deployed as an interactive web application using Streamlit Community Cloud. The application is designed to be simple, fast, and accessible to non-technical users.
- Supports single customer prediction
- Supports batch prediction using CSV upload
- User-friendly interface with clear outputs
- Live application available online
- https://ecommerce-churn-prediction-3nwr22bnvx9mwus6wcp562.streamlit.app/

KEY LEARNINGS

- This project provided hands-on experience in building a complete machine learning solution from data to deployment.
- End-to-end ML pipeline development
- Handling real-world data challenges
- Feature engineering and model evaluation
- Cloud deployment and dependency management

FUTURE IMPROVEMENTS

- The current system can be enhanced further to increase scalability and business value.
- Real-time data integration
- Automated model retraining
- Model explainability using SHAP
- Integration with marketing automation tools

THANK YOU

Thank you for your time and attention.