NAAN MUDHALVAN

1105-Gojan school of business and Technology (GSBT)

Course Name: Data Analytics in Process Industries

PROJECT TITLE

Customer conversion analysis for online shopping using click stream data

PRESENTED BY:

S.Santhiya -110523237023

S.Selvarani- 110523237025

S.Shivani keerthana -110523237026

S.Vaithiswari-110523237029

B.Varshini-110523237030

Customer conversion analysis for online shopping using click stream data

Abstract:

In the digital commerce era, understanding customer behavior is vital for optimizing sales. Click stream data—records of users' interactions with websites—offers deep insights into online shopping behaviors. This study explores how analyzing click stream data can enhance customer conversion rates by identifying patterns, bottlenecks, and optimal engagement strategies. Using data mining, user session analysis, and machine learning, this approach enables e-commerce platforms to make data-driven decisions that improve user experience and drive sales.

Problem Statement:

As a Data Scientist at an e-commerce giant (like Amazon, Walmart, or eBay), your task is to develop a Streamlit-based intelligent web application that uses clickstream data to:

- **Classification:** Predict whether a customer will make a purchase based on browsing behavior.
- **Regression:** Estimate potential revenue generated by individual customers.
- **Clustering:** Segment users into groups to enable targeted marketing and personalized experiences.

This solution helps businesses enhance customer engagement, increase conversions, and optimize marketing strategies.

Introduction:

Customer conversion is the process of turning casual website visitors into paying customers. In e-commerce, click stream data serves as a digital footprint of user activity. Analysing this data provides insights into user behavior, preferences, and decision-making paths. This study focuses on mining click stream data to understand user intent, identify drop-off points, and implement strategies to boost conversions.

Domain

- ❖ E-commerce and Retail Analytics

Customer conversion analysis for online shopping using click stream data

Business Use Cases:

- Customer Conversion Prediction
- Revenue Forecasting
- Customer Segmentation
- Churn Reduction
- Personalized Product Recommendations

Goal:

The primary goal is to analyze click stream data to identify behavioral patterns and features that influence customer conversion. Secondary objectives include detecting drop-off points, segmenting user types, and recommending website improvements to enhance user engagement and purchase likelihood.

Methodology

- **Data Collection**: Obtain click stream logs from an e-commerce website.
- **Preprocessing**: Clean and structure the data to remove noise and standardize sessions.
- **Session Identification**: Group click data into individual user sessions.
- **Feature Extraction**: Derive meaningful features such as time spent, product views, clicks before conversion, etc.
- **Modeling**: Use classification or clustering techniques (e.g., decision trees, logistic regression, k-means) to predict conversion likelihood.
- **Evaluation**: Assess model accuracy using metrics like precision, recall, F1-score, or AUC.

Approach Overview:

1. **Data Preprocessing:**

- **Handle Missing Values**: Imputed using mean/median/mode.
- **Feature Encoding**: One-Hot/Label Encoding.
- **Scaling**: MinMaxScaler or StandardScaler.

Customer conversion analysis for online shopping using click stream data

2. **Exploratory Data Analysis (EDA) :**

- Visualize trends using **bar charts**, **pair plots**, **heatmaps**.
- Session-based insights: **duration**, **page views**, **bounce rates**.
- Temporal features: **Hour**, **Weekday**, **Browsing patterns**.

3. **Feature Engineering:**

- Session metrics: Click counts, session time.
- Behavioral patterns: Bounce, revisit, exit rates.
- Sequential patterns in clickstreams.

4. **Handling Imbalance (Classification Only) :**

- Techniques used: **SMOTE**, **Random Undersampling**, **Class Weighting**.

5. **Model Building:**

- **Classification Models**: Logistic Regression, Decision Trees, Random Forest, XGBoost, Neural Networks.
- **Regression Models**: Linear, Ridge, Lasso, Gradient Boosting Regressors.
- **Clustering Models**: K-Means, DBSCAN, Hierarchical Clustering.
- **Pipelines**: Used Scikit-learn pipelines for full automation.

6. **Model Evaluation :**

- **Classification**: Accuracy, Precision, Recall, F1, ROC-AUC
- **Regression**: MAE, MSE, RMSE, R^2
- **Clustering**: Silhouette Score, Davies-Bouldin Index

7. **Streamlit Application :**

- Interactive interface for:
  - CSV file upload / manual input
  - Conversion prediction (classification)
  - Revenue estimation (regression)
  - Cluster visualization (customer segments)
  - Real-time plots and analytics

Customer conversion analysis for online shopping using click stream data

Advantages

- Enables personalized marketing and product recommendations.
- Improves customer journey mapping.
- Helps reduce bounce rates and abandoned carts.
- Allows real-time tracking and intervention.
- Drives data-informed business decisions.

Disadvantages

- High volume of data can be complex to manage and analyze.
- Privacy concerns regarding user tracking.
- Requires significant computational resources.
- Interpretability of machine learning models can be challenging.
- Noise and inconsistency in raw click data can affect accuracy.

Project Outcomes:

- Achieved **high accuracy and precision** in customer conversion prediction.
- Enabled **data-driven revenue estimation** based on user activity.
- Provided **actionable customer segments** to guide marketing efforts.
- Delivered a **user-friendly Streamlit app** with dynamic visualizations and model results.

Evaluation Metrics:

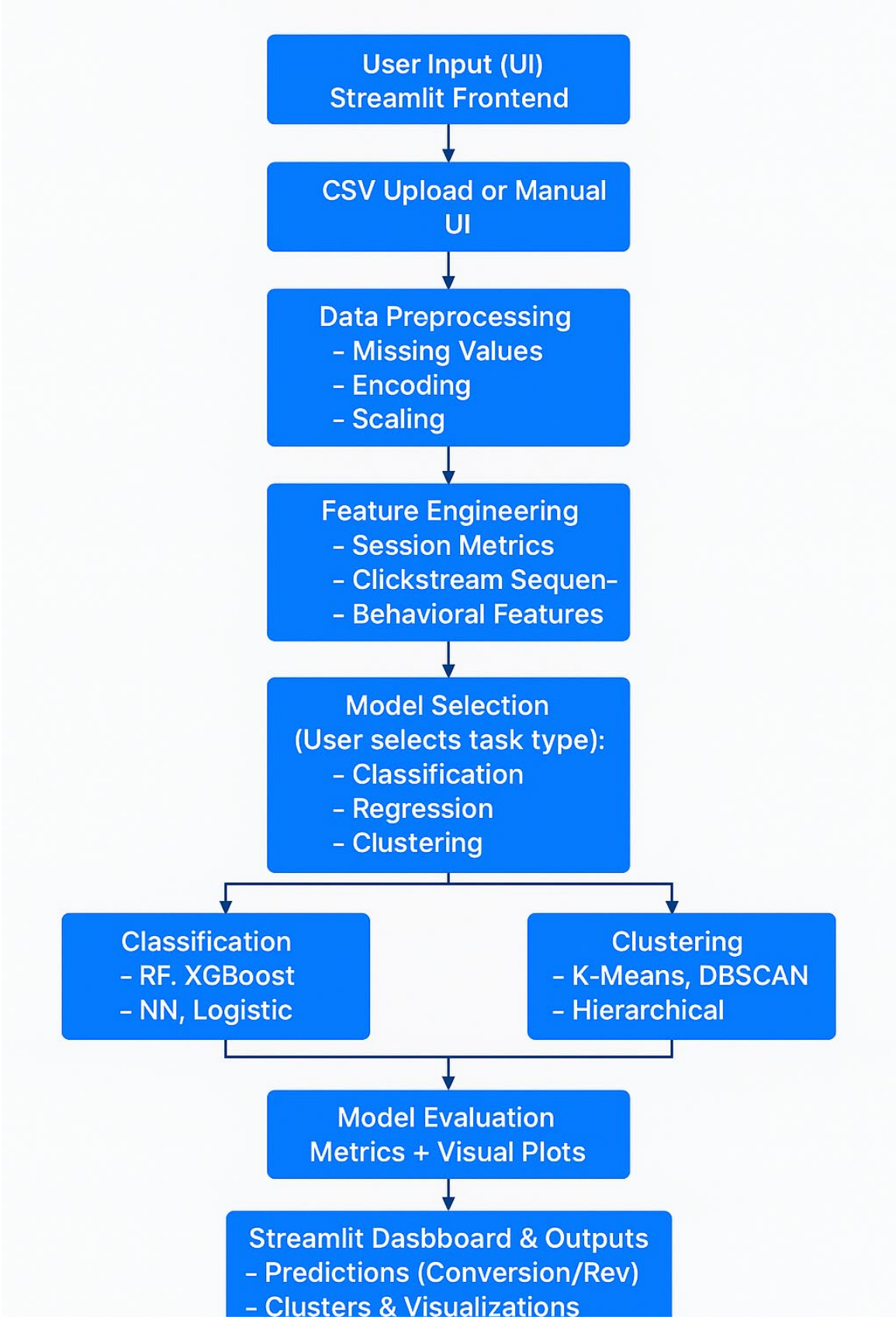
Task	Metrics Used
Classification	Accuracy, Precision, Recall, F1-Score, ROC-AUC
Regression	RMSE, MAE, R-squared
Clustering	Silhouette Score, Davies-Bouldin Index

Skills & Tools Gained:

- **Languages & Libraries**: Python, Pandas, NumPy, Matplotlib, Seaborn
- **ML Frameworks**: Scikit-learn, XGBoost, TensorFlow
- **Techniques**: Data Preprocessing, Feature Engineering, Hyperparameter Tuning
- **Visualization**: Streamlit, EDA Tools
- **Deployment**: Interactive web app for model inference and insights

Customer conversion analysis for online shopping using click stream data

System Architecture:



Customer conversion analysis for online shopping using click stream data

Motivation

With increasing competition in online markets, understanding why users do or do not convert is crucial for business success. Traditional analytics often fall short in capturing the full picture of user interactions. Click stream analysis allows businesses to uncover detailed behavior trends, optimize user journeys, and ultimately increase revenue through improved conversion rates.

Conclusion

Click stream data analysis offers a powerful approach to understanding and improving customer conversion in online shopping. By mining and modeling user behavior, businesses can uncover critical insights into what drives users to purchase—or not. Despite challenges in data processing and privacy, the potential benefits for e-commerce optimization are substantial.