**Customer Conversion Analysis for Online Shopping Using Clickstream Data**

**1. Introduction**

This project aims to analyze online shopping behavior using clickstream data to predict customer conversions, estimate revenue, and segment customers. The solution is implemented using machine learning techniques and deployed as an interactive Streamlit web application.

**2. Methodology**

**a. Data Preprocessing**

* **Handling Missing Values**: Median imputation for numerical features, mode for categorical features.
* **Feature Encoding**: Label Encoding for categorical variables.
* **Feature Scaling**: StandardScaler applied to numerical features.
* **Balancing Data**: SMOTE used for handling class imbalance in classification.

**b. Exploratory Data Analysis (EDA)**

* **Visualizations**: Histograms, bar charts, pie charts.
* **Session Analysis**: Page views, session duration, and bounce rate.
* **Correlation Analysis**: Heatmaps to identify feature relationships.

**c. Machine Learning Models**

* **Classification**: RandomForestClassifier to predict if a customer will complete a purchase.
* **Regression**: GradientBoostingRegressor to estimate potential revenue.
* **Clustering**: KMeans clustering to segment customers based on browsing behavior.

**d. Model Evaluation**

* **Classification Metrics**: Accuracy, Precision, Recall, F1-score.
* **Regression Metrics**: MAE, RMSE, R-squared.
* **Clustering Metrics**: Silhouette Score.

**3. Streamlit Application**

The web application provides real-time predictions for:

* **Conversion Probability**
* **Estimated Revenue**
* **Customer Segmentation**

**Features:**

* **CSV Upload**: Allows bulk predictions.
* **Manual Data Entry**: Users can input data for real-time insights.
* **Interactive Visualizations**: Charts and graphs to interpret insights.

**4. Results**

* Achieved high accuracy in customer conversion prediction.
* Provided revenue estimation to help businesses optimize marketing strategies.
* Segmented customers effectively for targeted recommendations.

**5. Deployment**

* The Streamlit app can be deployed on **AWS Elastic Beanstalk, EC2, or Streamlit Cloud**.
* Uses **MLflow** for experiment tracking and model management.

**6. Conclusion**

This project provides an end-to-end machine learning solution for e-commerce businesses, helping them understand customer behavior, predict conversions, and optimize revenue.