Machine Algorithm Activities

Activity 1: Segment customers based on spending patterns across product categories.

Instructions:

1. **Data Collection:** Create a dataset with customer spending in categories like electronics, groceries, and clothing.
2. **Implement K-Means Clustering:** Group customers into 2 or 3 clusters using KMeans.
3. **Cluster Analysis:** Interpret the results.
4. **Evaluation Metrics:**
   - **Silhouette Score:** Measures how similar a sample is to its own cluster compared to others.

Activity 2: Detect fraudulent transactions from a dataset containing transaction data.

Instructions:

1. **Dataset Utilisation:** Use a dataset with transaction features like amount and type.
2. **Data Exploration and Preprocessing:** Address class imbalance and clean the data.
3. **Data Splitting:** Divide the dataset into training and testing sets.
4. **Model Development:** Build a binary classification model, such as Random Forest or XGBoost.
5. **Evaluation Metrics:**
   - **Confusion Matrix:** Shows the number of correct and incorrect predictions.
   - **Accuracy:** Measures the proportion of correct predictions.
   - **Precision:** Evaluates the accuracy of positive class predictions.
   - **Recall:** Measures the ability to detect all actual positive instances.
   - **F1 Score:** A balance between precision and recall.
   - **ROC-AUC Score:** Assesses how well the model distinguishes between classes.

Activity 3: Predict Customer Lifetime Value (CLV) based on purchase behaviour and other features.

Instructions:

1. **Dataset Exploration:** Understand the dataset's features related to CLV.
2. **Data Cleaning and Encoding:** Handle missing data and encode categorical features.
3. **Data Splitting:** Split the data into training and testing sets.
4. **Regression Model Development:** Use a regression model like Linear Regression to predict CLV.
5. **Evaluation Metrics:**
   - **Mean Squared Error (MSE):** Measures the average squared difference between predicted and actual values.
   - **Root Mean Squared Error (RMSE):** The square root of MSE, indicating the model's prediction error.

Showcase your work from all the activities to your peers/mentor and discuss.