-:Programming Assignment:-

Objective:

This assignment requires the implementation of different classification algorithms on the given dataset.

Dataset:

The dataset captures the purchasing behavior of online shoppers, featuring 17 attributes that describe various aspects of their shopping sessions. Each session is labeled as either positive (where a purchase was made) or negative (no purchase occurred). Your task is to classify these sessions based on the provided features, distinguishing between successful and unsuccessful shopping experiences.

[Online Shoppers Purchasing Intention Dataset](#)

Algorithm Research:

Study the following classification algorithms:

- K-Nearest Neighbour (KNN)
- Logistic Regression
- Random Forest
- Gradient Boost
- AdaBoost
- Support Vector Machine (SVM)

Feel free to use any other algorithm of your choice also.

Implementation:

- Use Python to implement the above given algorithms on the dataset.

- Utilize libraries such as scikit-learn for implementation.

Key Steps:

- Split the data into training and testing sets.
- Train each model on the training data.
- Make predictions on the test data.
- Calculate and compare classification accuracy for each algorithm.

Deliverables:

1. **Jupyter/Google Collab Notebook** containing:
   - Data loading and preprocessing
   - Implementation of all the given classification algorithms
   - Model training and prediction
   - Evaluation metrics (focus on classification accuracy)
2. **Brief Report** (2-3 pages) summarizing:
   - Dataset and its characteristics
   - Brief explanation of implemented algorithms
   - Comparison of classification accuracies
   - Insights on algorithm performance

Brownie Points:

There is inherent class imbalance in the dataset, with 84.5% of sessions belonging to the negative class (no purchase) and only 15.5% to the positive class (purchase). Try to identify and implement an algorithm that is resilient to this imbalance, ensuring accurate classification despite the skewed distribution.