**Class 11: Classification through ML Models from SVM to XGboost**

**Instructor: Ansar Rahman**

**Assignment**

**1. Dataset Exploration and Preprocessing (20%)**

* Download Titanic Dataset and Iris Dataset from Kaggle.
* Perform exploratory data analysis (EDA):
  + Summarize the dataset (e.g., data types, missing values, unique classes, and descriptive statistics).
* Preprocess the dataset:
  + Handle missing data if any.
  + Encode categorical variables using suitable methods (e.g., one-hot or label encoding if required).
  + Normalize / Standardize or scale features if required.

**2. Model Implementation (20%)**

* **Train and Evaluate Models**:
  + Implement the following classifiers:
    - Support Vector Machine (SVM)
    - Decision Tree
    - Random Forest
    - XGBoost
  + Split the data into training and testing sets.
  + Use cross-validation to fine-tune hyperparameters for each classifier.
* **Hyperparameter Tuning**:
  + Experiment with different hyperparameters for each model. Examples:
    - SVM: Kernel types (linear, poly, RBF).
    - Decision Tree: Max depth, criterion (gini vs entropy).
    - Random Forest: Number of estimators, max depth.
    - XGBoost: Learning rate, max depth, n\_estimators.

**3. Model Evaluation and Comparison (20%)**

* Evaluate each model's performance using:
  + Accuracy
  + Precision, Recall, and F1-score
* Generate confusion matrices for each classifier.
* Compare model performance and discuss their strengths and weaknesses for the chosen dataset.

**4. Reporting and Visualization (20%)**

* Create visualizations to compare the models, such as:
  + Bar plots for evaluation metrics (e.g., accuracy, F1-score).
* Summarize the findings in a report:
  + Which model performed best and why?
  + How does the choice of hyperparameters impact performance?

**Good Luck 🙂**