Assignment 3

The main objective of this assignment is to provide hands-on experience in implementing and understanding Support Vector Machines (SVMs) and Neural Networks for classification tasks. Each team will work together to solve a classification problem using both models, compare their results, and analyze the performance.

Team Setup:

- Each team will consist of 3 members.
- Team members will collaborate to implement the models, prepare the data, and analyze the results.

1. Problem Definition:

Each team will use a publicly available dataset for classification. Some examples of datasets include:

- **Iris Dataset** (for flower classification)
- **MNIST Dataset** (handwritten digit classification)
- **Wine Quality Dataset** (classification of wine quality)
- **Breast Cancer Dataset** (classification of benign or malignant tumors)

Teams should choose one of these datasets or suggest an alternative dataset with instructor approval.

2. Data Preprocessing

- Load the dataset and check for missing values, and handle them (imputation, removal).
- Split the dataset into training and testing sets (70% training, 30% testing).
- Normalize/standardize the dataset if necessary.
- Perform exploratory data analysis (EDA) to understand the data better and create visualizations (e.g., histograms, pair plots).

3. Support Vector Machine (SVM) Implementation

- Implement a Support Vector Machine using a popular machine learning library (e.g., sklearn in Python).
- Tune the SVM by experimenting with different kernels (linear, polynomial, RBF).
- Train the SVM on the training data and evaluate its performance on the testing data.
- Report the classification accuracy and confusion matrix.

4. Neural Network Implementation

- Implement a simple neural network using libraries like Keras or PyTorch.
- Use a basic feed-forward network with at least one hidden layer.
- Experiment with different activation functions (ReLU, Sigmoid, etc.).
- Train the network on the training data and evaluate its performance on the testing data.
- Report the classification accuracy, confusion matrix, and loss curve.

5. Model Comparison and Analysis

- Compare the performance of the SVM and the Neural Network on the test set.
- Use performance metrics such as accuracy, precision, recall, and F1-score.
- Create visualizations (e.g., ROC curves) to compare the models' effectiveness.
- Analyze the results, identify any overfitting or underfitting, and suggest improvements.

Required files to be submitted:

- **Code**: Submit the code for both models (SVM and Neural Network) along with any necessary documentation (e.g., setup instructions, dependencies).
- **Report**: Submit the report as a PDF document or in a shared format (Google Docs, etc.).