Machine Learning

*4 Phase Task Details:*

1. **1st Phase:**

* Apply necessary data cleaning and preprocessing steps to prepare the dataset for model training.

1. **2nd Phase:**

* *Tasks:*

1. Apply data preprocessing

2. Train Logistic Regression model on the train set

3. Test the trained model on the test set

* *Evaluate the performance on the test set using the following metrics:*
* Accuracy
* Confusion Matrix
* Precision
* Recall
* F1 Score

5. *Plot the following learning curves:*

* Accuracy (y-axis) vs Solver (x-axis)
* Accuracy (y-axis) vs Max\_iter (x-axis)
* Two models for two targets-
* Injury Type
* Patient Status

Note:

Solver = {'lbfgs', 'liblinear', 'newton\_cg', 'newton-cholesky', 'sag', 'saga'}

Max\_iter = {50, 100, 150, 200, 250, 300}

1. **3rd Phase:**

* Tasks:
* Apply necessary data pre-processing
* Train the following models on the train set:
* Decision Tree
* Random Forest
* XGBoost
* Support Vector Machine (SVM)
* Evaluate the performance of the trained models on the test set using the following metrics:
* Accuracy
* Precision
* F1 Score
* Recall
* Confusion Matrix
* Prepare a comparison table for the models' performance, including Logistic Regression from Phase 2.
* Plot the following curves:
* Decision Tree: Accuracy (y-axis) VS Max\_depth (x-axis)
* SVM: Accuracy (y-axis) VS Kernel (x-axis)
* Random Forest: Accuracy (y-axis) VS N\_estimators (x-axis)
* XGBoost: Accuracy (y-axis) VS Learning\_rate (x-axis)

NOTE:

* For Decision Tree, Max\_depth = {4, 5, 6, 7, 8, 9}
* For SVM, Kernel = {linear, poly, rbf, sigmoid}
* For Random Forest, N\_estimators = {10, 50, 100, 200}
* For XGBoost, Learning\_rate = {0.001, 0.01, 0.1, 1}

1. **4th Phase:**

* Tasks:
* Apply necessary data pre-processing
* Train an Artificial Neural Network (ANN) on the training Dataset.
* Evaluate the performance of the trained models on the test set using the following metrics:
* Accuracy
* Precision
* F1 Score
* Recall
* Confusion Matrix
* Prepare a comparison table for all the models' performance, including Logistic Regression, Decision Tree, Random Forest, XGBoost, SVM and ANN.
* Show necessary plots to compare the algorithms' performance.