

Continuous Evaluation: 70%, Viva: 30%

Assignment 3:

- i. Download the [Forest Cover Type Dataset](#) and preprocess the dummy variables to create training, test and development set. Reduce the train data size if the system is unable to process the whole dataset.
 - a. Consider only two features and three classes and train Logistic Regression 3-class classifier (any three class) to show the training and test accuracy plot in a 2D plane, using matplotlib.
 - b. Analyze and control the overfitting by varying the inverse of regularization strength parameter (0.1, 0.25, 0.5, 0.75, 0.9) and plot the accuracy graph for the test set.
 - c. Apply multiclass classification in Support Vector Machine (SVM) using Forest Cover Type dataset.
 - d. Plot and analyze the confusion matrix for the above applied SVM method.
- ii. Download [Titanic Dataset](#) and do the initial preprocessing and train a Decision Tree classifier and vary the maximum depth of the tree to train at least 5 classifiers to analyze the effectiveness.
 - a. Estimate the average accuracy of the Naïve Bayes Classifier using 5-fold cross-validation using a scikit-learn package in python. Plot the bar graph using matplotlib.

Submit a report with the result.