**DSE 2242– Fundamentals of Machine Learning Lab**

**Week 7 – Date: 14th February 2025**

**EXER 1**

* + - 1. Use the German credit rating dataset “German Credit Data.csv” , Decision tree classifier to predict good or bad credit. Use “sklearn.model\_selection” and GridSearchCV to search the hyperparameter values and report the most optimal one. Configure the grid search to search for optimal parameters:
      * Splitting criteria: gini or entropy.
      * Maximum depth of decision tree ranging from 2 to 10.
      * The searching of optimal parameter will be validated using 10-fold cross validation and the most optimal parameter will be chosen based on ROC AUC score.

2. Visualize the tree using graphviz toolkit.

3. Display the text representation of the rules learnt.

**EXER 2**

Use the iris dataset to implement , Decision tree classifier

Implement a decision tree classifier, Train it on 80% of the dataset and test on the remaining 20%.

Print the classification report and confusion matrix.

**Optimize Hyperparameters Using Grid Search**

* Use GridSearchCV to tune max\_depth, min\_samples\_split, and criterion.
* Report the best hyperparameter combination and corresponding accuracy.

**Pruning the Decision Tree**

* Implement post-pruning using cost-complexity pruning (ccp\_alpha).
* Find the optimal ccp\_alpha by plotting accuracy vs. ccp\_alpha values.

**Feature Importance Analysis**

* Extract and plot feature importance values from the trained decision tree.
* Drop the least important feature and retrain the model. Does accuracy improve or decline?