

**JADAVPUR UNIVERSITY**

**Faculty of Engineering & Technology**

**B.E. Information Technology - Fourth Year**

**SESSION: 2021 - 2022**

**SUBJECT NAME: MACHINE LEARNING LAB**

**SUBJECT CODE: IT/PC/B/S/411**

**FULL MARKS: 30**

**TIME: 3 HRS**

**Answer any THREE questions**

**(10x3=30)**

1. Apply different types of following machine learning models: (CO1) (10)
- Support Vector Machine (SVM)
  - Decision Tree
  - Random Forest
  - Naive Bayes

And compare as well as discuss the performances in terms of Accuracy, Precision, Recall and F-measure on the following datasets:

- Wine Dataset: <https://archive.ics.uci.edu/ml/datasets/wine>
- Ionosphere Dataset: <https://archive.ics.uci.edu/ml/datasets/Ionosphere>

Generate the respective confusion matrices with class labels (heat map).

2. Implement an ANN based model for classification task on the following datasets: (CO2) (10)

- Iris plants dataset: <https://archive.ics.uci.edu/ml/datasets/Iris/>
- Diabetes dataset: <https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html>
- Wisconsin Breast Cancer Dataset: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))

3. Apply different Hidden Markov Models (HMMs): GaussianHMM, GMMHMM and MultinomialHMM upon the following datasets and compare their performances (Accuracy, Precision, Recall, F-score) in tabular form: (CO3) (10)

- Iris plants dataset: <https://archive.ics.uci.edu/ml/datasets/Iris/>
- Diabetes dataset: <https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html>
- Ionosphere Dataset: <https://archive.ics.uci.edu/ml/datasets/Ionosphere>

4. Train and test any one of the following Deep Learning based models: VGG16 or ResNet50. Display the confusion matrices with class labels (heat map), Accuracy, Precision, Recall and F-measure on the following datasets: (CO3) (10)

- a. SAVEE: <http://kahlan.eps.surrey.ac.uk/savee/Download.html>
- b. EmoDB: <http://www.emodb.bilderbar.info/navi.html>

Generate the figures of Receiver Operating Characteristic (ROC) curve and Area Under Curve (AUC) for each dataset wise evaluation.

5. Apply any two of the comparisons for the clustering task: (CO4) (10)

- a. K-means versus, K-medoids/PAM,
- b. Dendrogram versus AGNES versus BIRCH
- c. DBSCAN versus OPTICS

on the **Wine Dataset**: <https://archive.ics.uci.edu/ml/datasets/wine> and use the following performance measures

- a. Silhouette Coefficient
- b. Calinski-Harabasz Index
- c. Davies-Bouldin Index

6. Implement both Reinforcement Learning and Deep Reinforcement Learning for finding the shortest path in any user-input graph. Compare the performance between RL and DRL in tabular representation. (CO5) (10)

**CO1: Discuss** the tools and/or libraries for artificial neural network (ANN) and **practice** simple examples using them

**CO2: Develop/Use** feed forward ANN to solve a given problem and **show** the effect of increase in hidden layers, back propagation etc.

**CO3: Construct** Hidden Markov Model with ANN, and CNNs to **solve** problems in signal processing.

**CO4: Construct** various machine learning algorithms to cluster different data sets and **discuss** their comparative performances.

**CO5: Manipulate** Reinforcement Learning tools/libraries to tackle interesting problems.