SLOT – L31+32

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CYCLE SHEET – I – WINTER SEMESTER 2021-2022

Programme Name & Branch: B.Tech (CSE)

Course Name: DATAWAREHOUSE AND DATAMINING LAB

Course Code: CSE3030

Source Code with Output screen in VTOP and MS Team

CYCLE SHEET -1

A. Load the Dataset file into WEKA. This Dataset consists of set of attributes and a class label. Use this dataset and apply the following data pre-processing transformations.

1. Determine, How many instances and attributes contained in the dataset?

2. Give the class label and number of instances in each class. Which class is dominant in the dataset?

3. Which Attribute has a sparse set of attributes from the mean? And Why?

4. Find how many outliers in the Dataset? What we should do on the outlier?

5. Create 3 Bin for first 2 attributes. Draw the Bar chart and the count for each?

6. Apply the Normalize for range [0, 3].Give mean value for each related attributes.

7. Discretise and normalize the dataset?

8. Apply J48 with default setting and determine the accuracy before and after data pre-processing.

B. Apply the following Classifiers over the dataset prepared by you with your own data covers the details of your academic records in the previous examinations.

Decision Tree

Naïve Bayes

K-NN

Linear Regressesion

CourseCode	CAT-1 (%)	CAT-2 (%)	FAT(%)	Grade
CSE3030	63	53	87	

Construct the Dataset with ARFF

Apply the required pre-processing suitable to apply the classifier

Consider the Grade column as a Class Label

Interpret the accuracy, precision, recall-measure of all the classifiers.

Compare the Classier and choose the best’s one.

^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^