Experiment-2

Aim: Explore weka Data Mining/Machine Learning Toolkit.

I. Downloading and/or installation of WEKA data mining toolkit.
II. Understanding features of WEKA toolkit such as Explorer, Knowledge flow interface, Experimenter, command-Line interface.
III. Navigate the options available in the WEKA (ex: select attributes panel, pre process Panel, classify panel, associate panel and visualize panel).
IV. Study the arff file format.
V. Explorer the available data sets in WEKA.
VI. Load a data set (ex: weather dataset, iris dataset, etc).
VII. Load each dataset and observe the following:
   i. List the attributes names and they types.
   ii. Number of records in each dataset.
   iii. identify the class attribute.
   iv. plot Histogram.
   v. Determine the number of records for each class.
   vi. Visualize the data in various dimensions.

Objective:

To understand the features of weka and to know about navigate options available in weka. And to load a data set and experiment the results of it.

Preprocessors:

The data that is collected from the field contains many unwanted things that lead to wrong details. Thus, the data must be preprocessed to meet the requirements of the type of analysis you are seeking. This is done in the preprocessing module.

Classifiers:

Classifiers in WEKA are the models for predicting nominal or numeric quantities. The learning schemes available in WEKA include decision trees and lists, instance-based classifiers, classifiers include bagging, boosting, stacking, error-correcting output codes are locally weighted learning

WEKA: wai kato environment for knowledge Analysis.

The WEKA GUI chooser provides a starting point for launching WEKA's main GUI applications and supporting tools. The GUI chooser consists of four buttons - one for each of the four major WEKA applications - and four menus.

Explorer: It is an environment for exploring data explorer consists of several tabs. They are

→ preprocess: It is the first step in machine learning is to process the data. It is used to select data file, process it and make it fit for applying the various machine learning algorithms.

→ Classify: The classify tab provides you several machine learning algorithms for the classification of your data such as linear Regression, logistic Regression.

→ Cluster: Under the cluster tab there are several clustering algorithms provided - such as simple means, Filtered cluster, Hierarchical cluster.

→ Associate: Under the Associate tab you would find Aprion filtered Associater and FD Growth.

→ Select Attributes: Select Attributes allows you feature selection based on several algorithms such as classifier, Subset Eval, principal components.

→ Visualize tab: The visualize option allows you to visualize your processed data for analysis.

Simple CLI:

It provides a simple command line interface and allows direct execution of WEKA commands.

Experimentor:

It is an environment for performing experiment and conduction statistical tests between learning schemes.

knowledge Flow:

It is a java-Beans based interface for setting up and running machine learning experiments.

trees J48 classifier:

It is an algorithm to generate a decision tree that is generated by C4.5. It is also known as statistical classifier. for decision tree classification, we need a database.

weather-Nominal:

In weka, attributes can be nominal or numeric. The value of a nominal attribute is represented by a word: sunny, overcast and rainy for the outlook attribute: yes and no for play attribute.

arff file format:

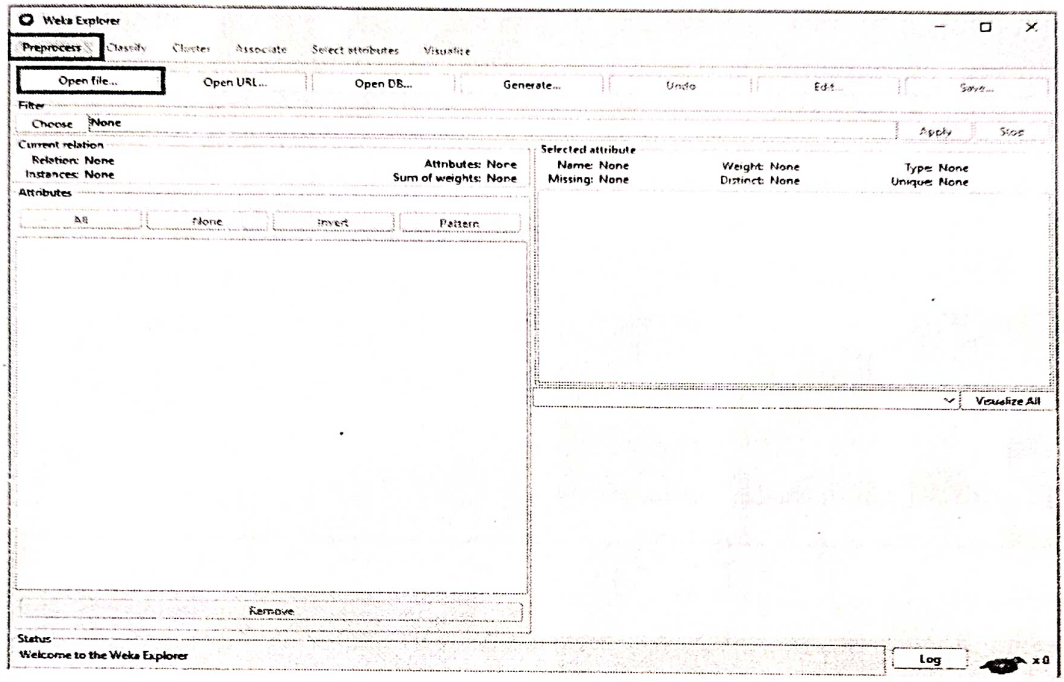
An ARFF (=Attribute-Relation File Format) file is an ASCII text file that describes a list of sharing a set of attributes. ARFF files are not the only format one can load, but all files that can be converted with WEKA's "core converters".

Procedure:-

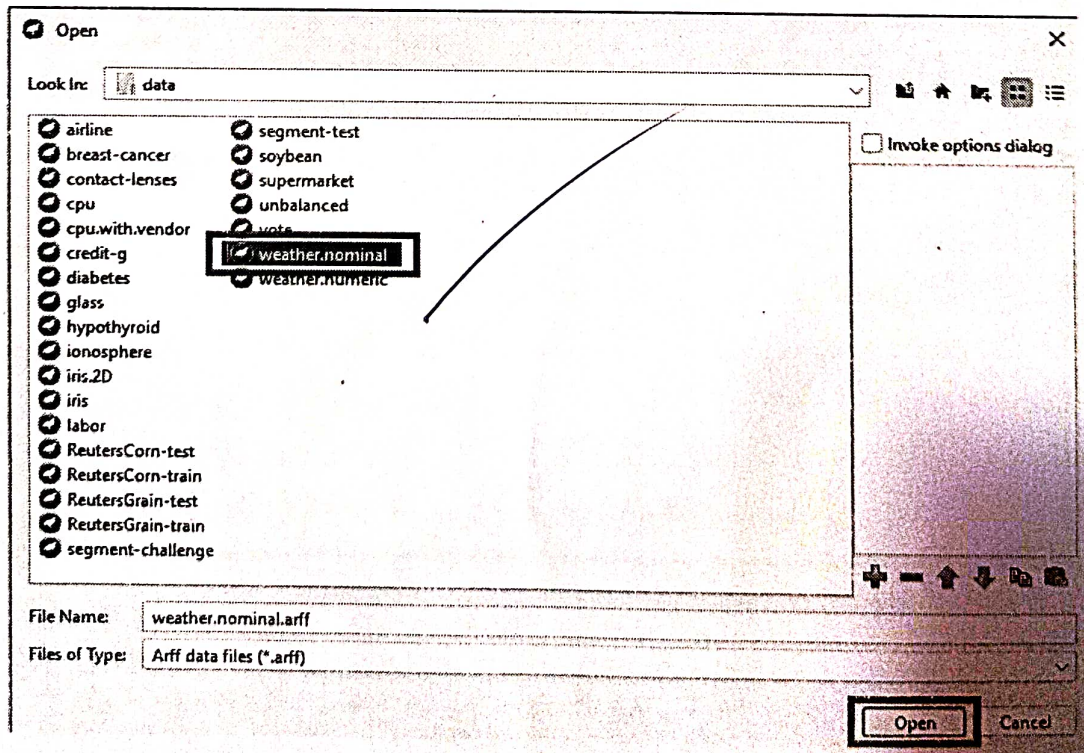
1) Open WEKA. you can see 5 tabs on the right side of the application, those are: Explorer, Experimentor, Knowledge Flow, workbench, Simple CLI.

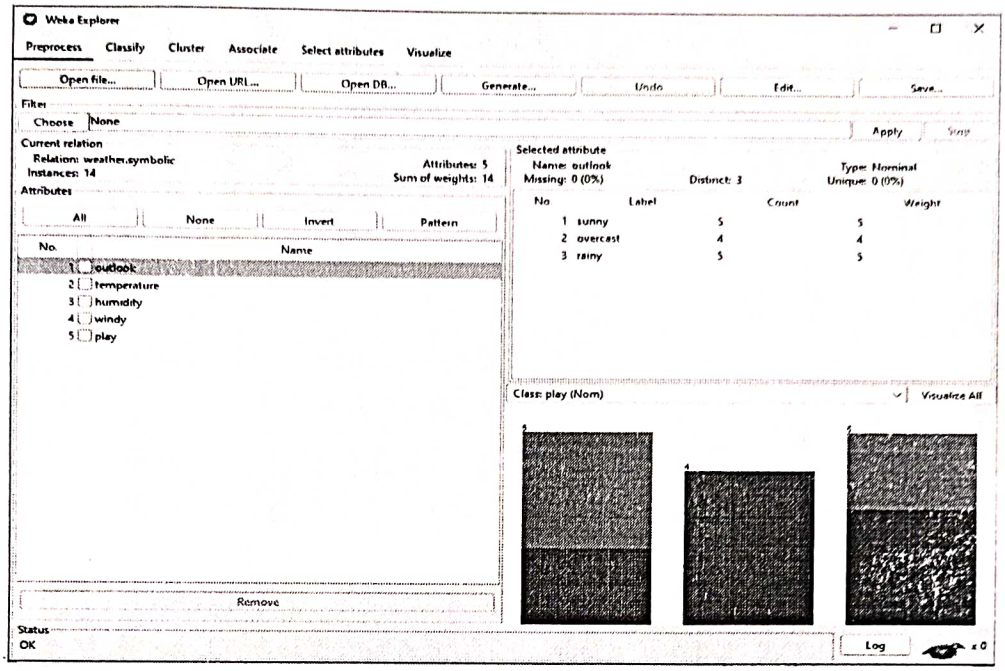
2) Click on Explorer.

3) On preprocess, click on "Open file".

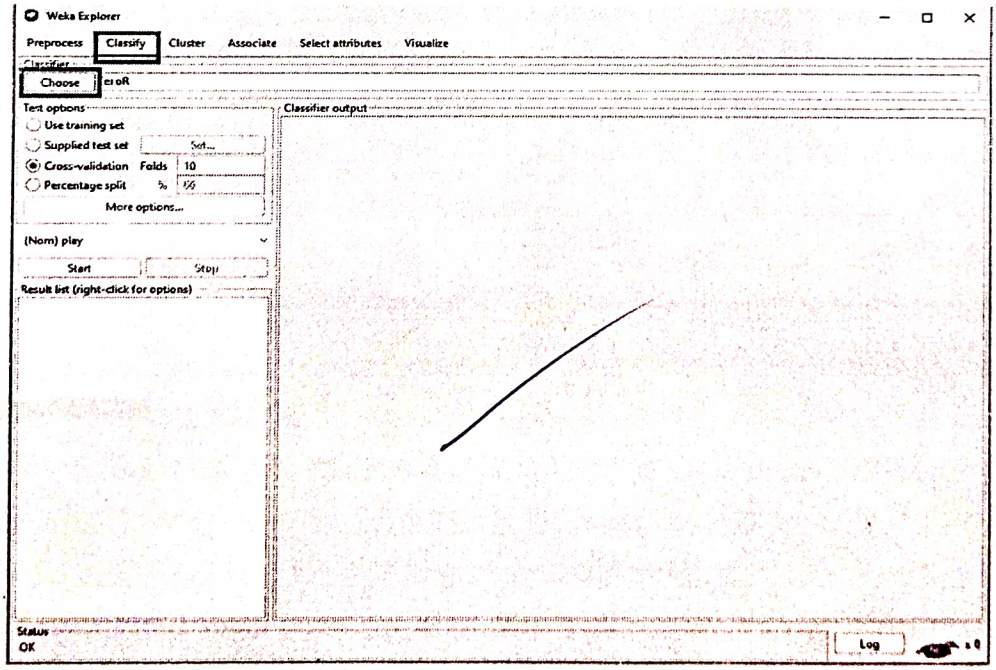


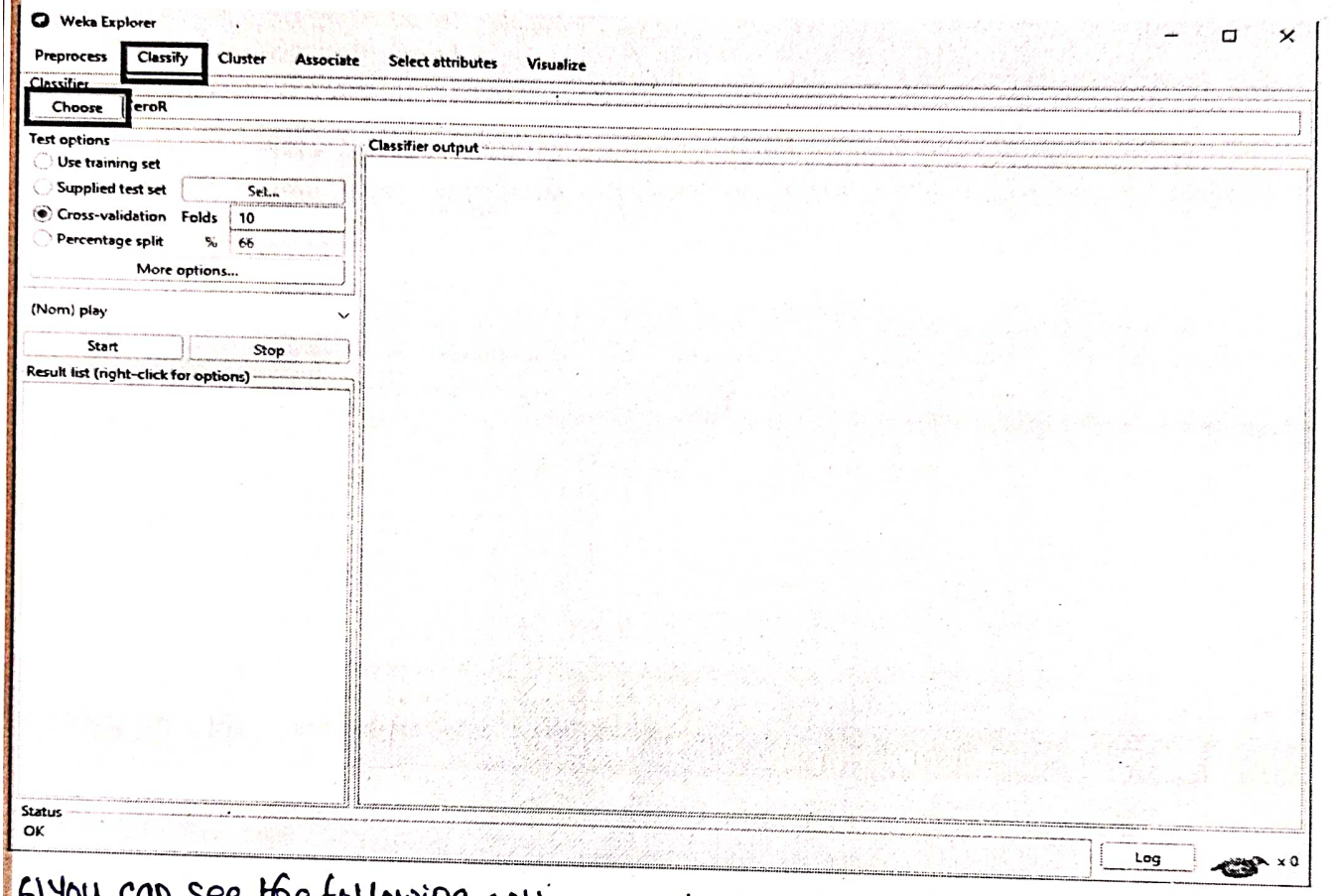
4) Go to "C:\Program Files\Weka-3-8-6\data". Select "weather.nominal.arff" and click on open.



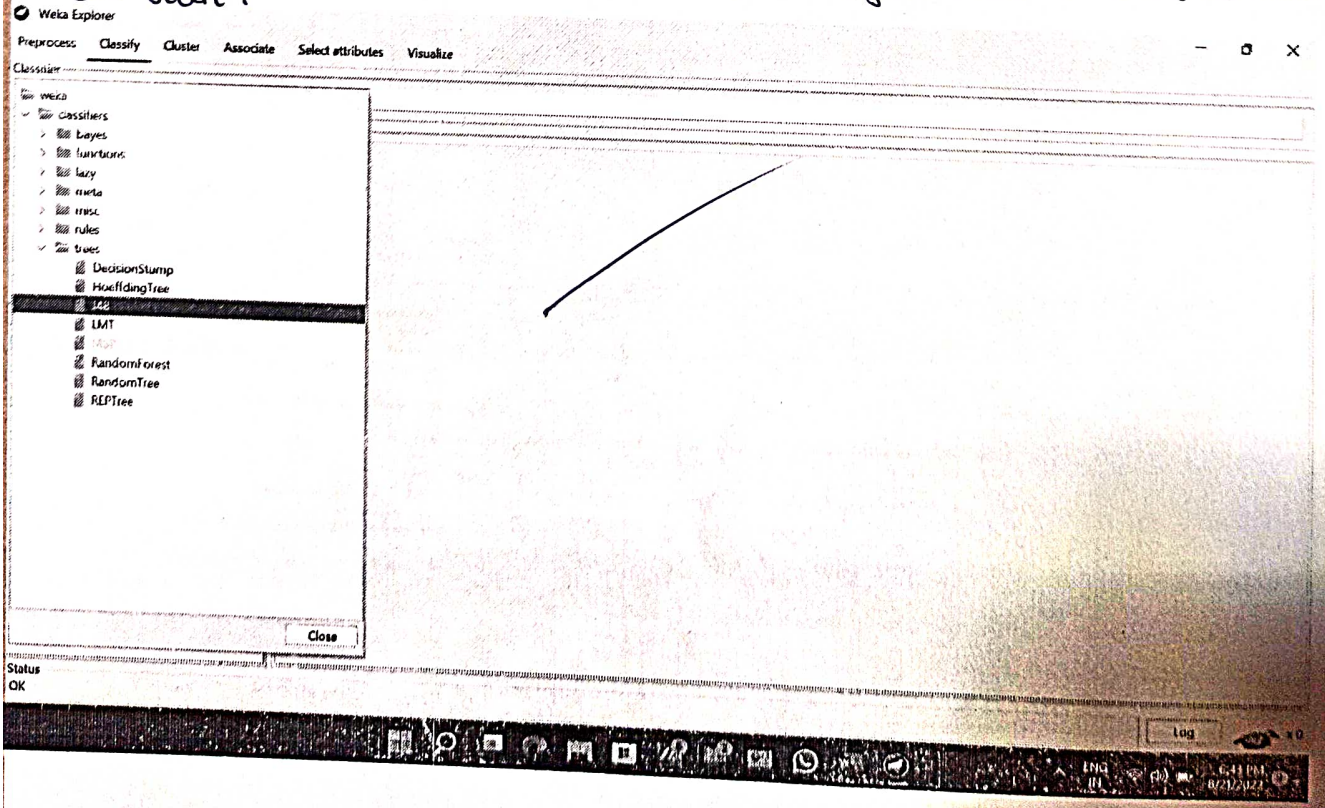


5) Click on "classify" as shown in below figure and then click on choose.

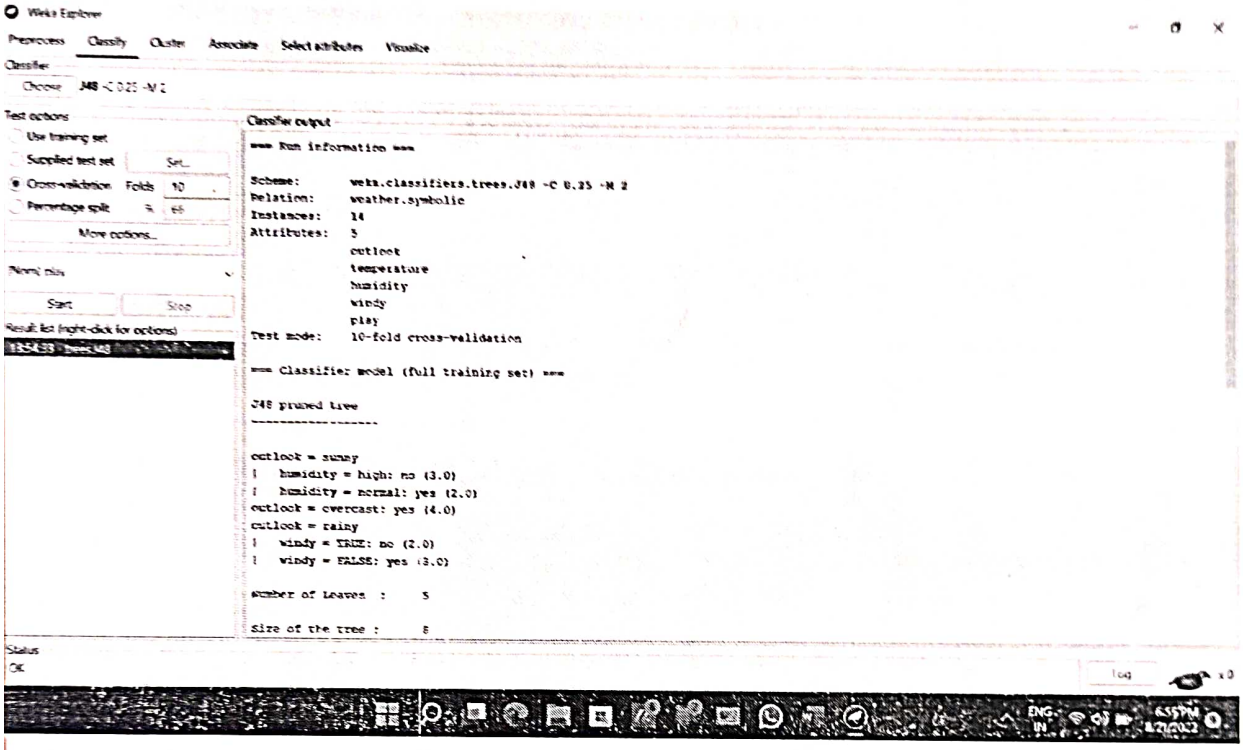




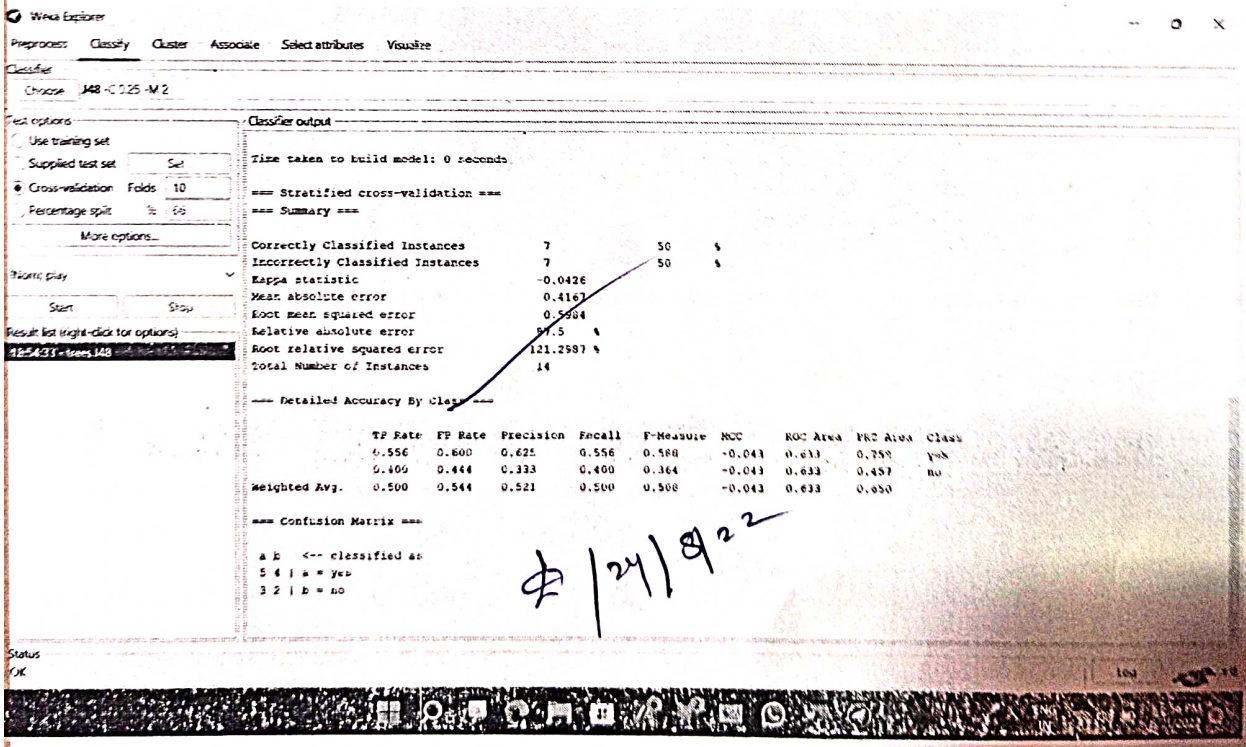
6) You can see the following options as shown in figure below select j48 and click on "Start".



7) You will see the following



8) Click on the resulted list to see the visual.



24/09/22

Output:

By the end of the experiment, we learned the installation of WEKA tool and we have understood the features of WEKA tool. and we have learned how to load a dataset in WEKA.

Experiment-4

Aim: Demonstrate performing classification on data.

→ Load each dataset into weka and run Id3, J48 classification algorithm. Study the classifier output. Compute the classifier output. Compute entropy values, kappa statistic.

→ Extract if-then rules from the decision tree generated by the classifier, observe the confusion matrix.

→ Load each dataset into weka and perform Naive-bays. classification and K-Nearest Neighbour classification. Interpret the results obtained.

→ Plot Roc curves.

→ Compare classification result of ID3, J48, Naive-Bayes and K-NN classifiers for each data set, and reduce which classifier is performing best and poor for each dataset and justifies.

Objective:

The ultimate objective of classification is to relate a variable of interest with observed variables. The actual variables of interest is meant to be of "Qualitative" type. The algorithm required for performing the classification is known as the classifier.

Zero-R:

→ ZeroR is the simplest classification method which relies on the target and ignores all predictors.

→ ZeroR classifier simply predicts the majority category.

→ Although there is no predictability power in ZeroR it is useful for determining a baseline performance as a benchmark for other classification methods.

One-R:

→ This method is used in the sequential learning algorithm for learning the values.

→ It returns a single rule that covers atleast some examples.

→ However, what makes it really powerful, is its ability to create relations among the attributes given. Hence covering a larger hypothesis space.

Explorer: It is an environment for exploring data.

Simple CLI: It provides a simple command-line interface and allows direct execution of WEKA commands.

Experimentor: It is an environment for performing experiment and conducting statistical tests between learning schemes.

Knowledge Flow: It is an Java-Beans based interface to setting up and running machine learning experiments.

Preprocess: It is the first step in machine learning to preprocess the data. It is used to select the data file preprocessing and make it fit for applying the various machine learning algorithms.

Classify: The classify tab provides you several machine learning algorithms for the classification of your data. Such as linear Regression, Logistic Regression.

Test options: Before you run the classification algorithm, you need to set test options. Set test options. Set test options in the Test options box. The test options that available are.

1. Use training set:

Evaluates the classifier on how well it predicts the class of the instances it was trained own.

2. Supplied test set:

Evaluates the classifier on how well it predicts the class of a set of instances it was trained on. Loaded from a file.

Clicking on the 'set...' button brings up a dialog allowing you to choose the file to test on.

3. Cross Validation:

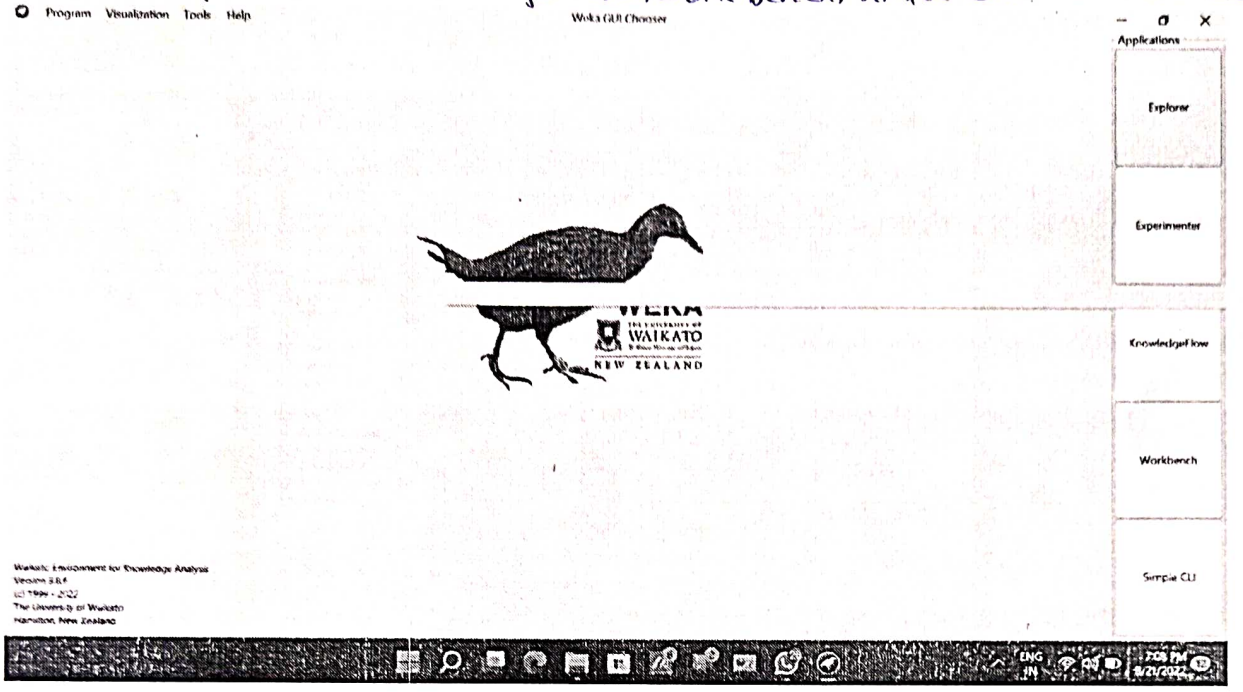
Evaluates the classifier by cross-validation, using the number of folds that are entered in the 'Folds' text field.

4. Percentage split:

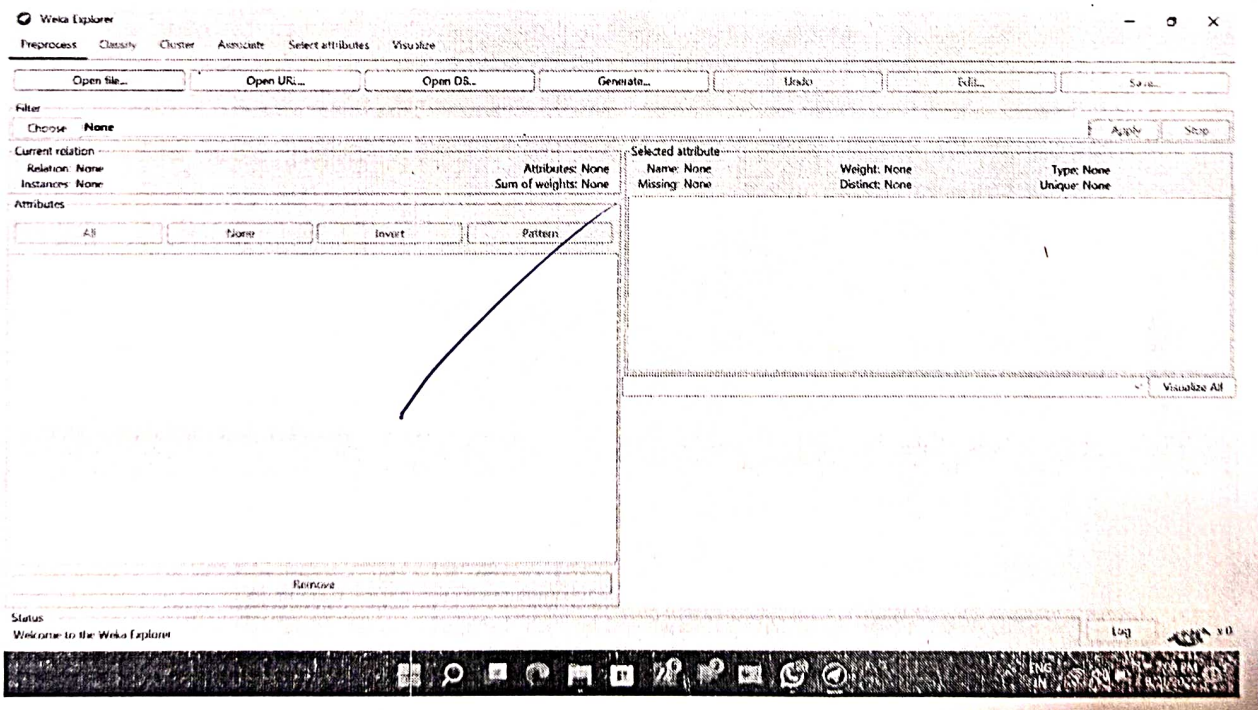
Evaluates the classifier on how toell it predicts a certain percentage of the data, which is held out for testing. The amount of data held out depends on the value entered in '%' field.

procedure:-

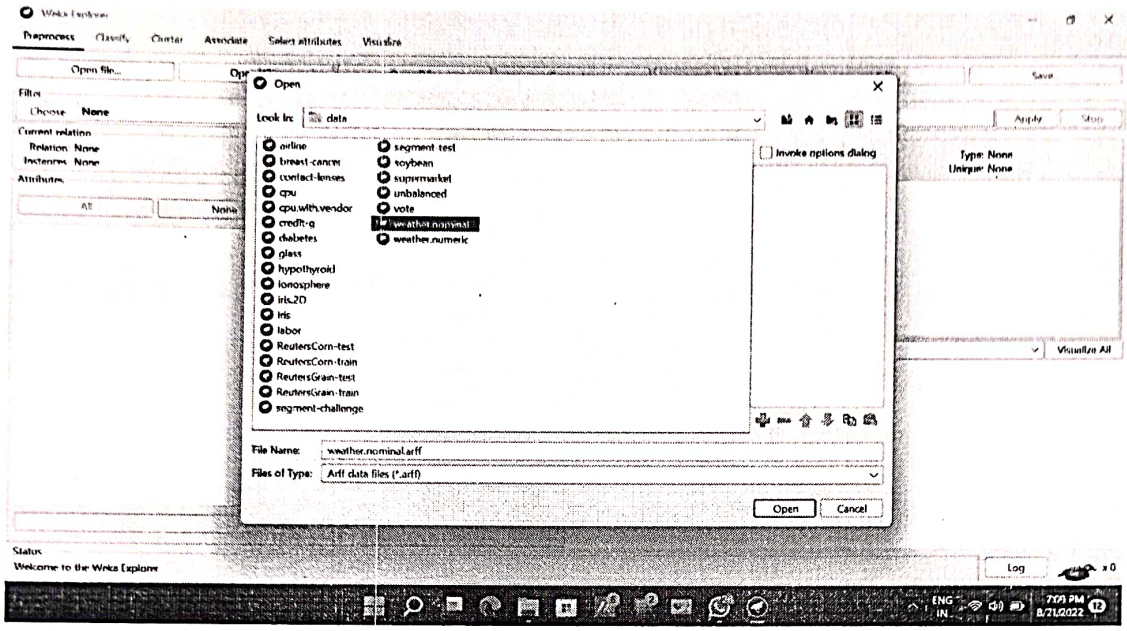
1) Open WEKA you can see 5 tabs on the right side of the application. Those are explorer, Experimentor, Knowledge Flow, work bench, simple CLI. Click on Explorer.



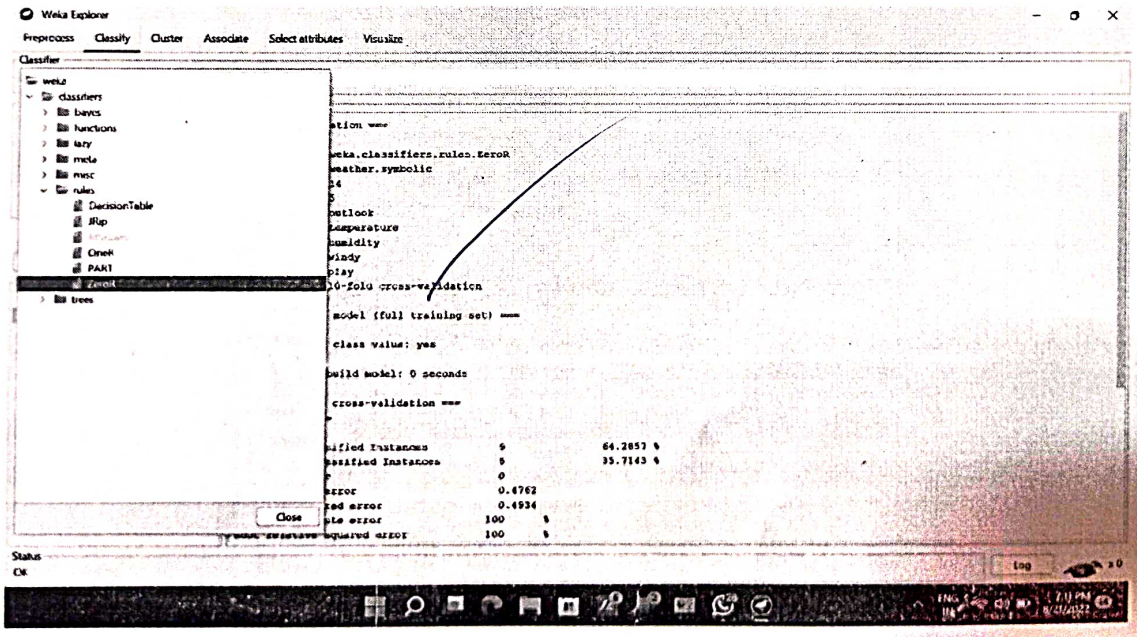
2) on preprocess. Click on "open file".



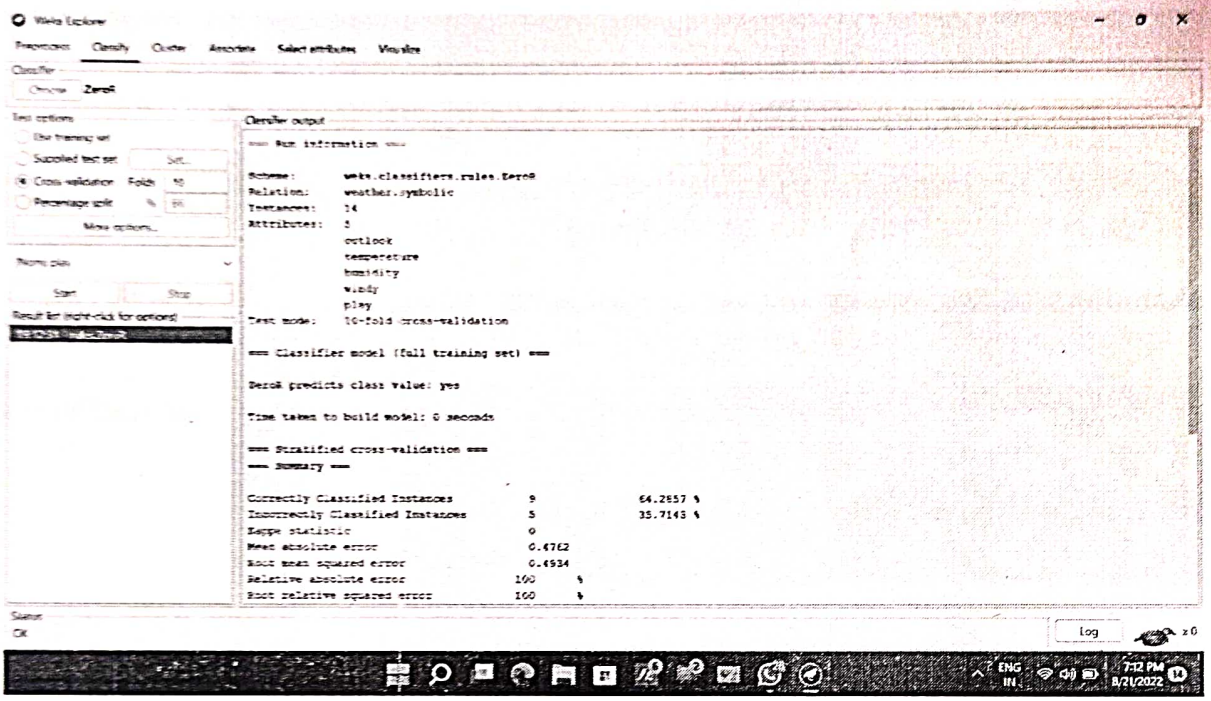
3) Go to "C:\Program Files\Weka-3-8-6\data", select "weather.nominal.arff" and click on open.



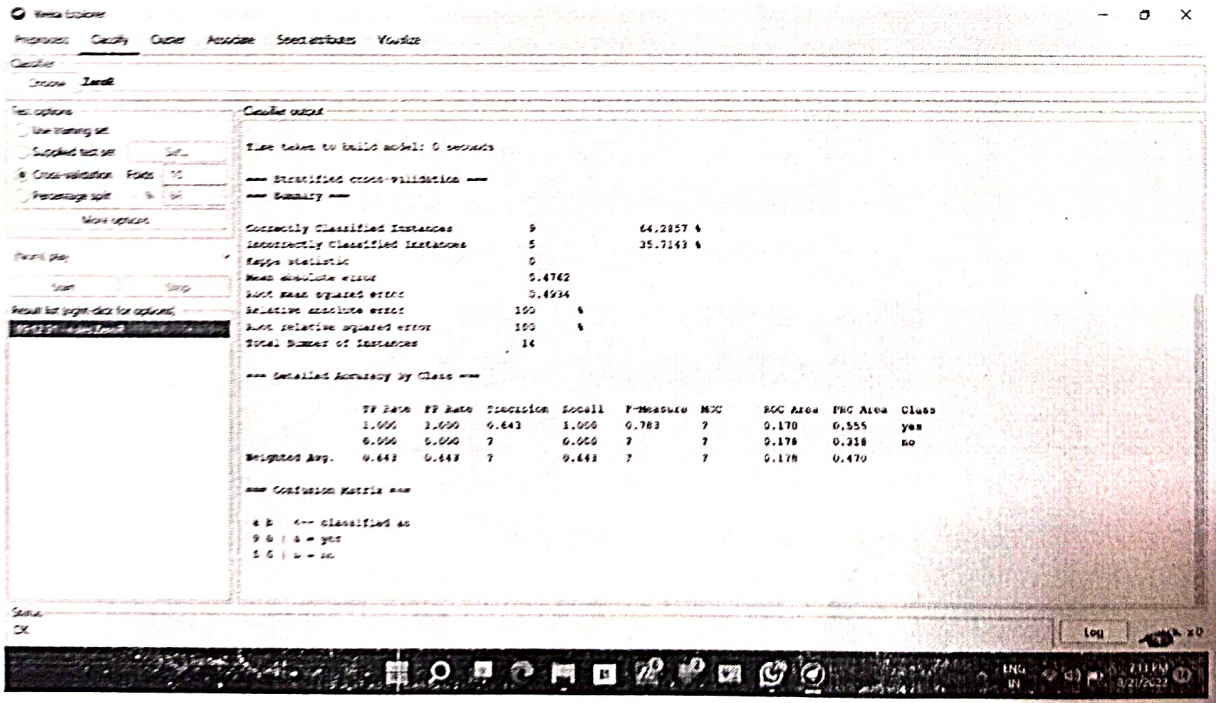
4) You can observe choose ZeroR. Click on it.



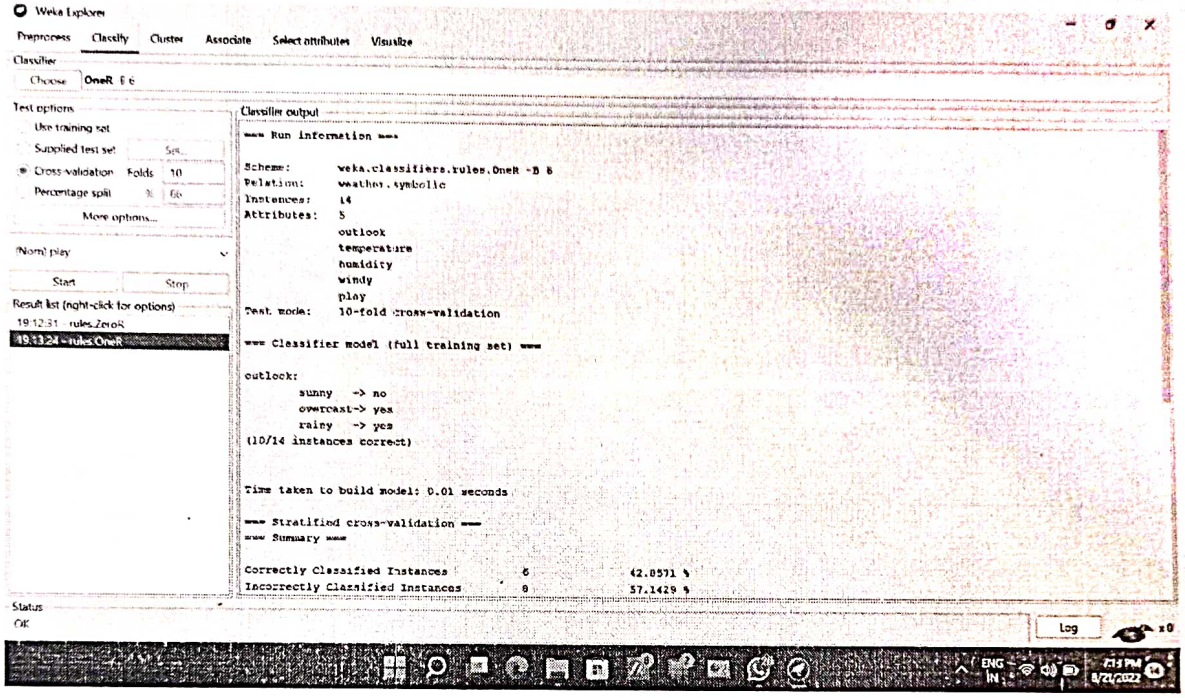
5) In test option you can see cross-validation folds. set it as 10. and click on "Start" button.

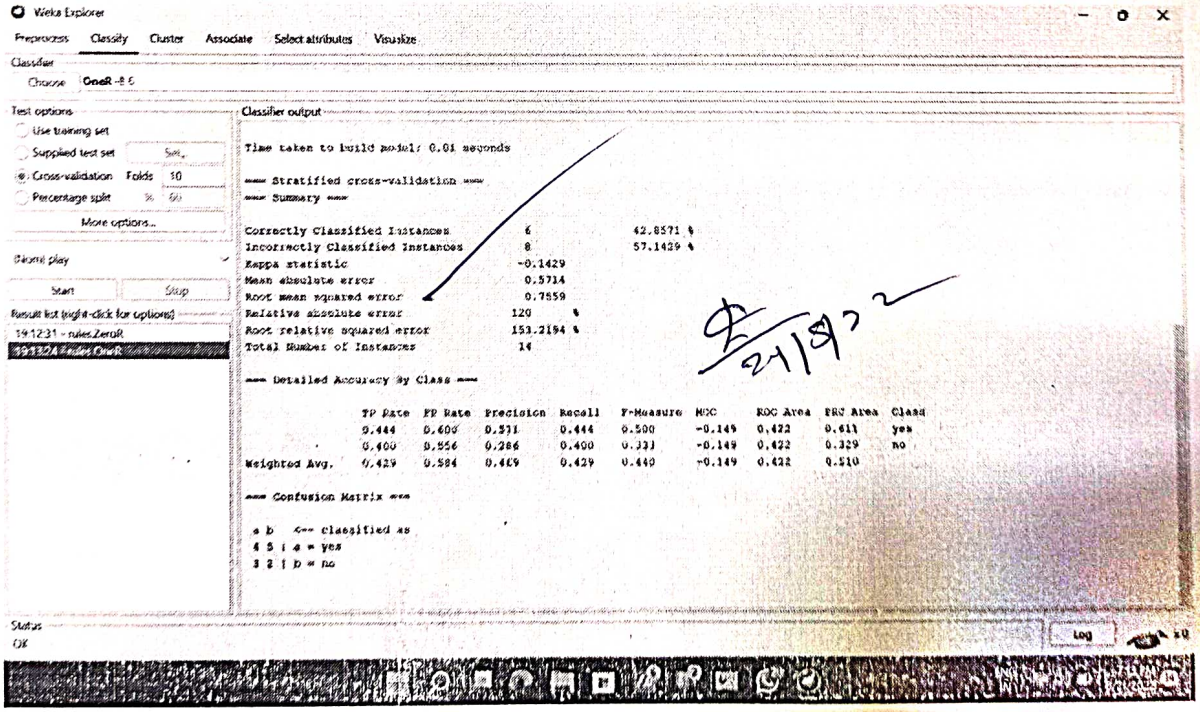


6) You can scroll down in the classifier output and see the mean absolute error, Root mean squared error etc.



7) Click on 'choose' and select 'OneR' and set cross validation folds to 10 and click on "start" button.





Output:

By the end of the experiment we have experimented with ZeroR, OneR algorithm and we observed the difference between these two algorithms.