|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Experiment No:** | 08 |  |  |  |  |  | TE AI&DS | |
| Date of Performance: | |  |  |  |  |  | Roll No: 9696 | |
| **Aim**: : To analyze and evaluate the performance of different Clustering techniques using WEKA tool | | | | | | | | |
| **Related CO5:** To analyze and evaluate perform of data mining techniques applied on large dataset using open-source tool for data mining | | | | | | | | |
| **Objective:**  To learn WEKA(Data Mining tool) and analyze and evaluate clustering algorithm performance. | | | | | | | | |
| **Rubrics for assessment of Experiment:** | | | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| **Sr. No** | **Parameters** | **Exceed** | |  | **Meet** | | **Below** | |
|  |  | **Expectations(EE)** | | | **Expectations** | | **Expectations** | |
|  |  |  |  |  | **(ME)** | | **(BE)** | |
| 1 | Timeline (2) | Early or on time | | | One session late | | More than one | |
|  |  | (2) |  |  | (1) |  | session late (0) | |
| 2 | Preparedness (2) | Knows the basic | | | Managed to | | Not aware of the | |
|  |  | theory related to | | | explain the | | theory to the | |
|  |  | the experiment | | | theory related to | | point. (1) | |
|  |  | very well. (2) | |  | the experiment. | |  |  |
|  |  |  |  |  | (1) |  |  |  |
| 3 | Effort (3) | Done expt on their | | | Done expt with | | Just managed. | |
|  |  | own. (3) | |  | help from other. | | (1) |  |
|  |  |  |  |  | (2) |  |  |  |
| 4 | Documentation(2) | Lab experiment is | | | Documented in | | Experiments not | |
|  |  | documented in | | | proper format | | written in | |
|  |  | proper format and | | | but some | | proper format | |
|  |  | maintained neatly. | | | formatting | | (0.5) |  |
|  |  | (2) |  |  | guidelines are | |  |  |
|  |  |  |  |  | missed. (1) | |  |  |
| 5 | Result (1) | Specific | |  | Partially | | Not specific at | |
|  |  | conclusion.(1) | | | specific | | all. (0) | |
|  |  |  |  |  | conclusion. | |  |  |
|  |  |  |  |  | (0.5) |  |  |  |
| **Assessment Marks:** | |  |  |  |  |  |  |  |
|  |  |  |  |  | |  | |  |
| Timeline(2) | Preparedness(2) | Effort(3) |  | Documentation(2) | | Result(1) | | Total(10) |
|  |  |  |  |  |  |  |  |  |

**Theory:**

WEKA is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset.

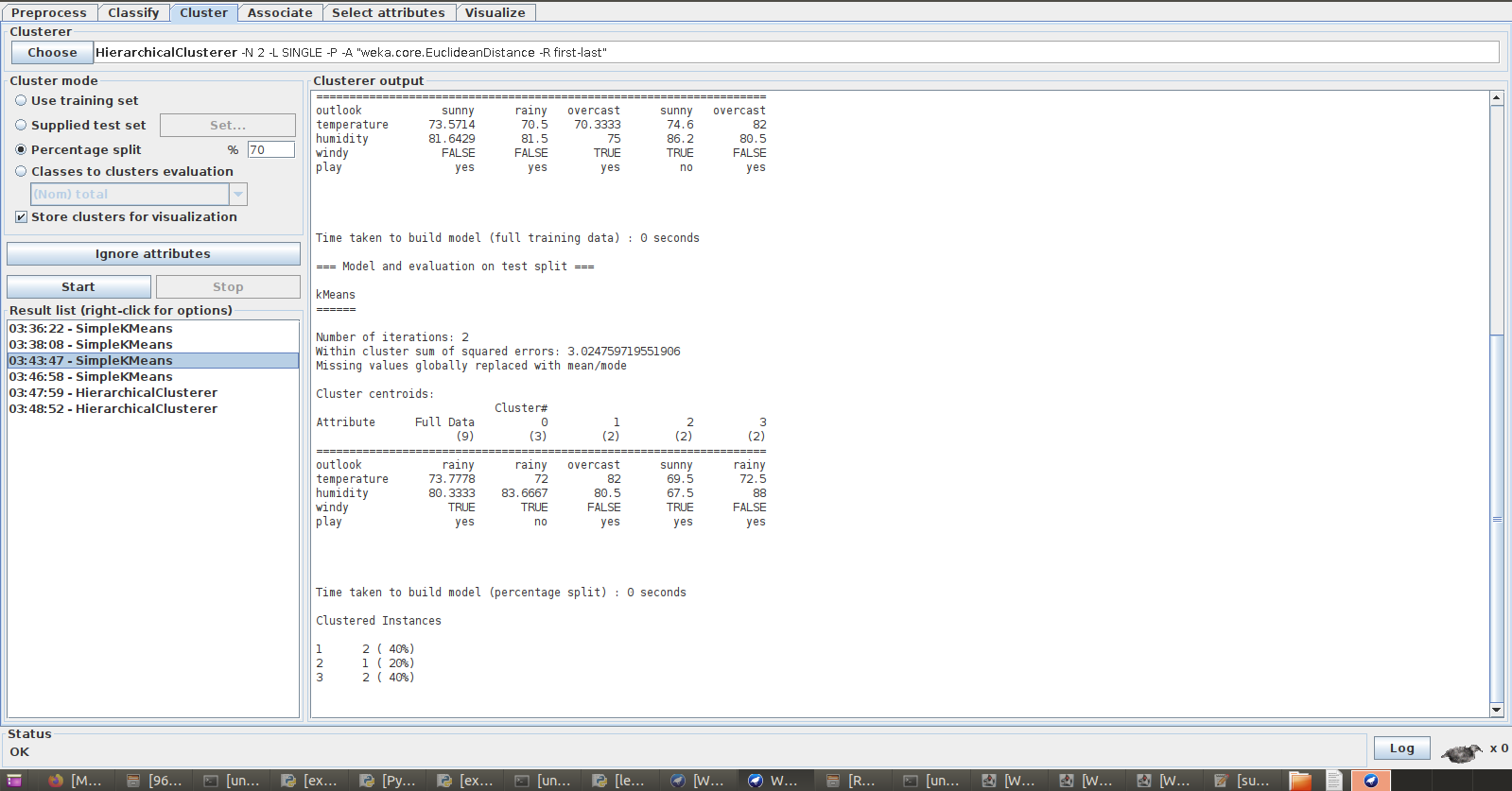
WEKA implements algorithms for data preprocessing, classification, regression, clustering, association rules; it also includes a visualization tools. When 'WEKA GUI Chooser' window appears on the screen, we can select one of the four options.

The options are 1. Simple CLI : provides a simple command line interface and allows direct execution of WEKA commands. 2. Explorer : is an environment for exploring data. 3. Experimenter : is an environment for performing experiments and conducting statistical tests between learning schemes. 4. Knowledge Flow : is a Java-Beans-based interface for setting up and running machine learning experiments. Classifiers in WEKA are the models for predicting nominal or numeric quantities. The learning schemes available in WEKA include decision trees, Bayes net, neural network, support vector machine and so on. In this experiment analysis and evaluation of three classification algorithm: Naive Bayesian algorithm, C4.5 algorithm, zero R is done. Before running the classification algorithm it is required to set test options. The selected test options were : 1. Use training set : Evaluates the classifier on how well it predicts the class of the instances it was trained on. 2. Cross-validation : Evaluates the classifier by cross-validation, using the number of folds (10) 3. Percentage split : Evaluates the classifier on how well it predicts a certain percentage of the data, which is held out for testing. In the classifier evaluation options following options are checked : 1. Output model : The output is the classification model on the full training set. 2. Output per-class stats : The precision/recall and true/false statistics for each class output 3. Output confusion matrix : The confusion matrix of one classifier's prediction is included in the output When training set is complete, the 'classifier' output area on the right panel of the classifier window is filled with text describing the results of training and testing.

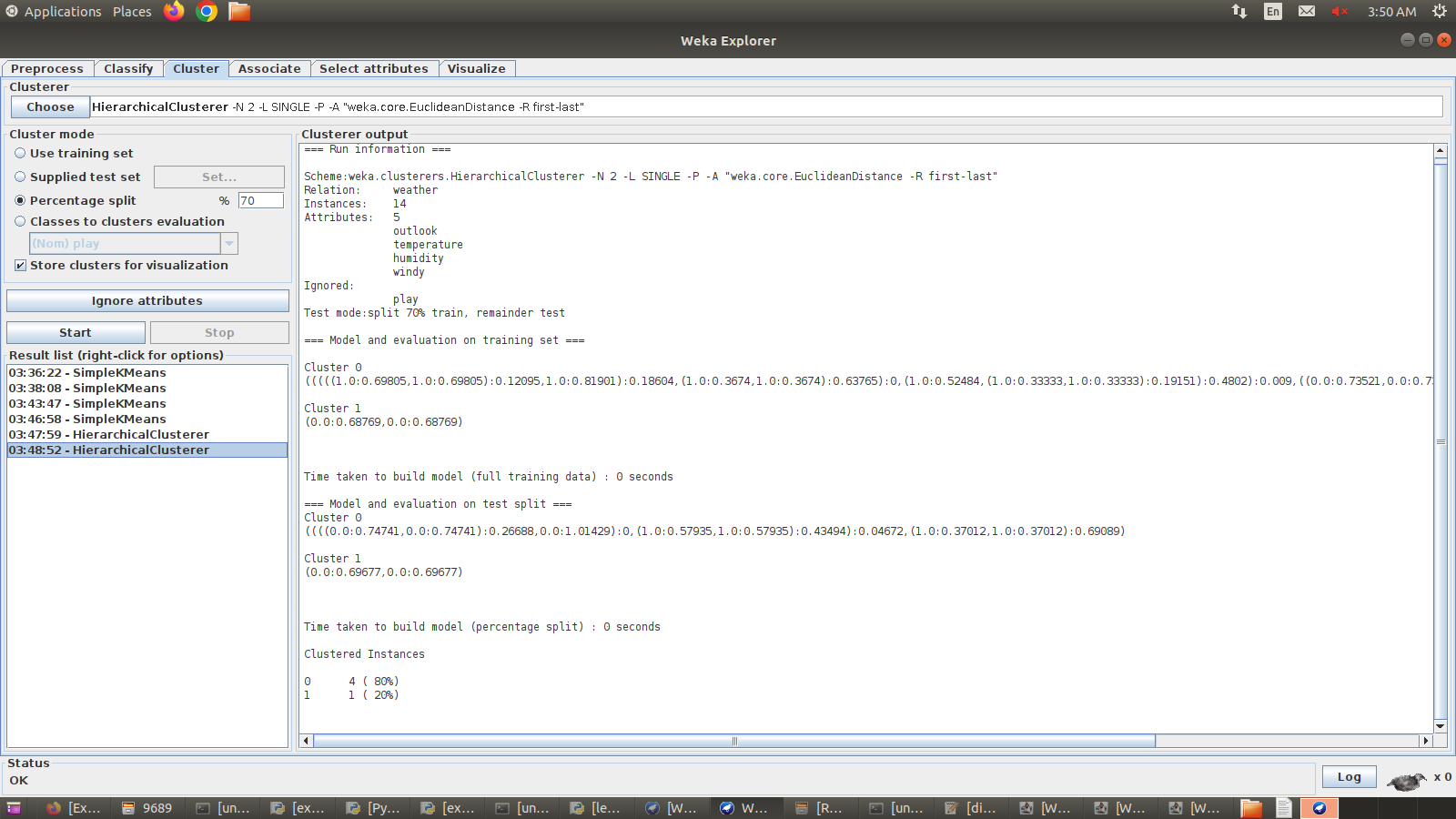
**Practical Exercise:**

Apply and evaluate the result for different classification techniques on various datasets using WEKA.

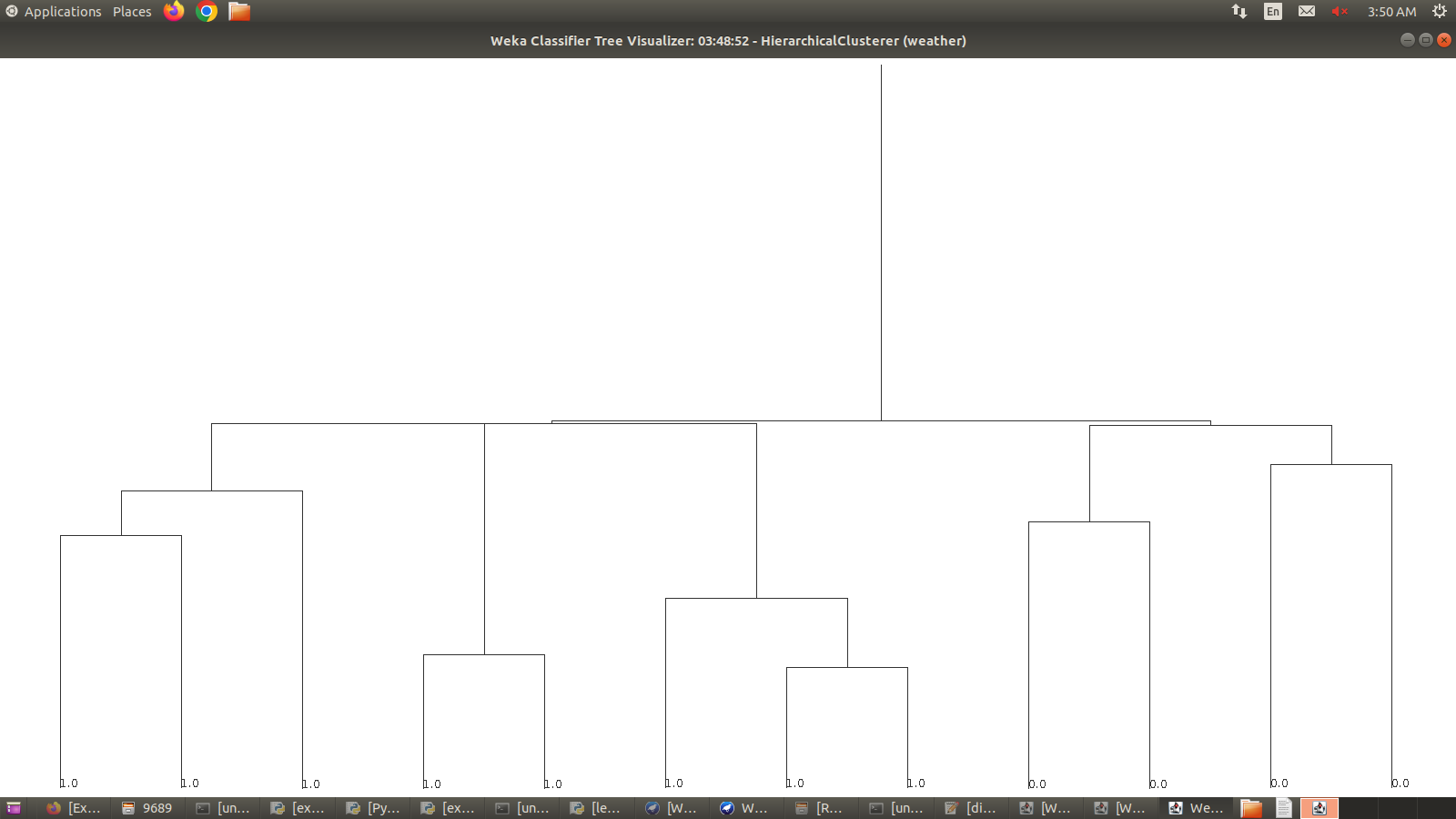
**K Mean Algorithm**

****

**Hierarchical Clustring:**

****

**Dendogram:**

****

**Post lab Questions:**

1. Apply agglomerative hierarchical clustering technique (single linkage) on following dataset.

A(1,2)

B(3,4)

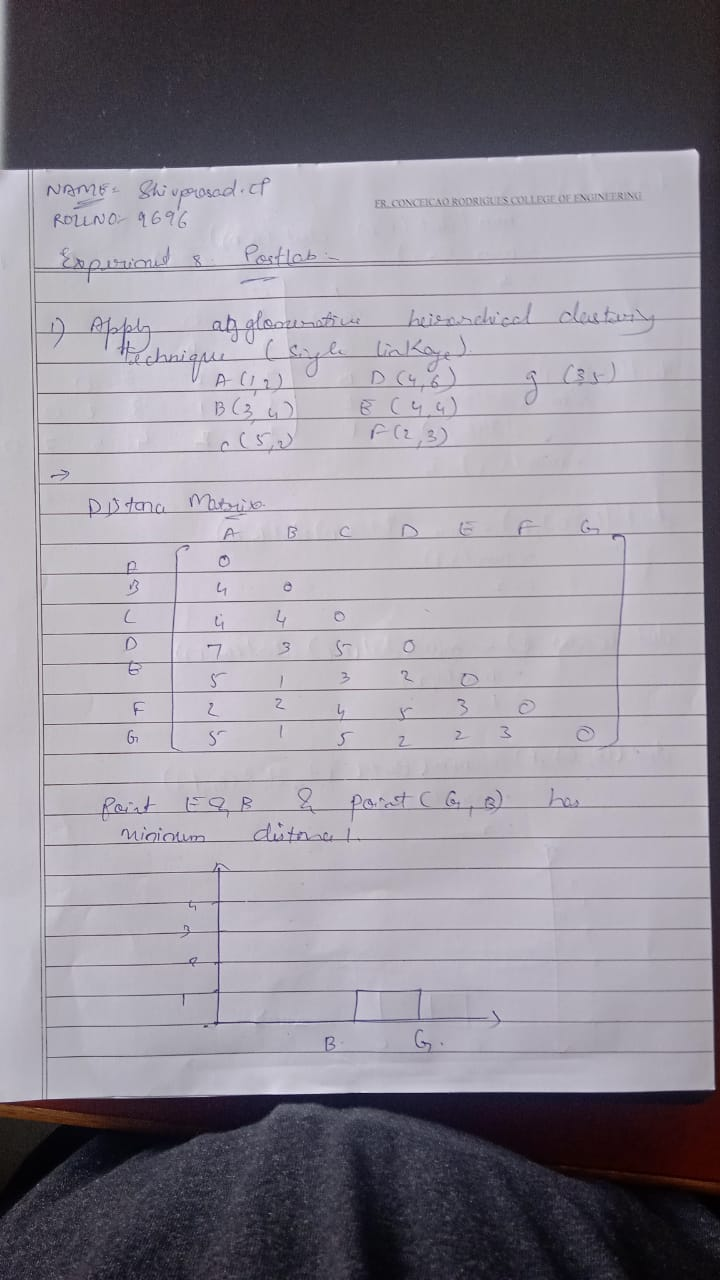
C(5,2)

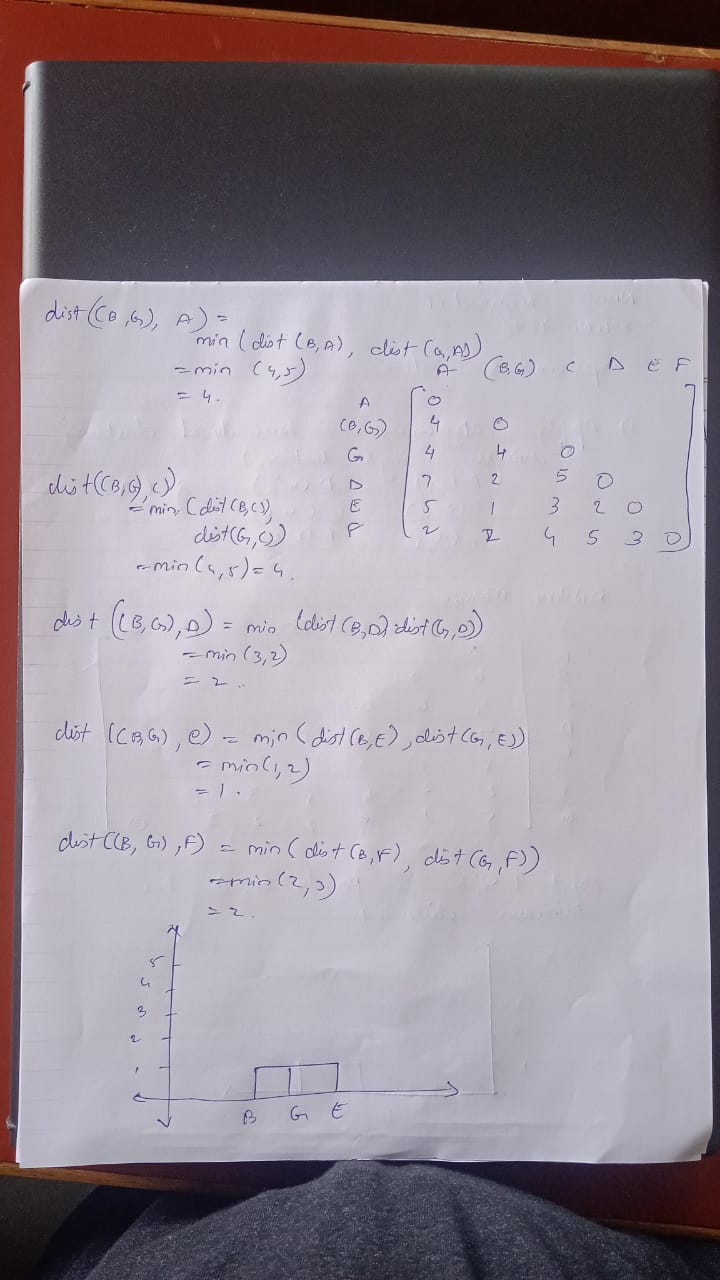
D(4,6)

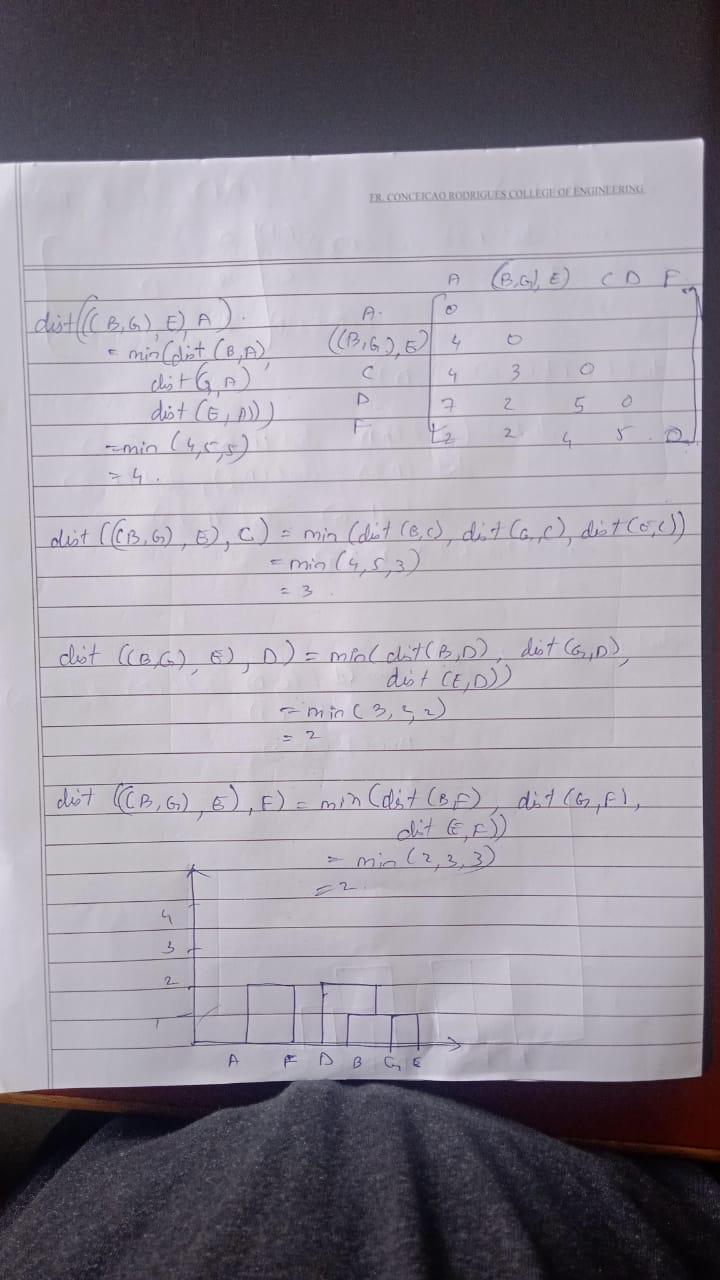
E(4,4)

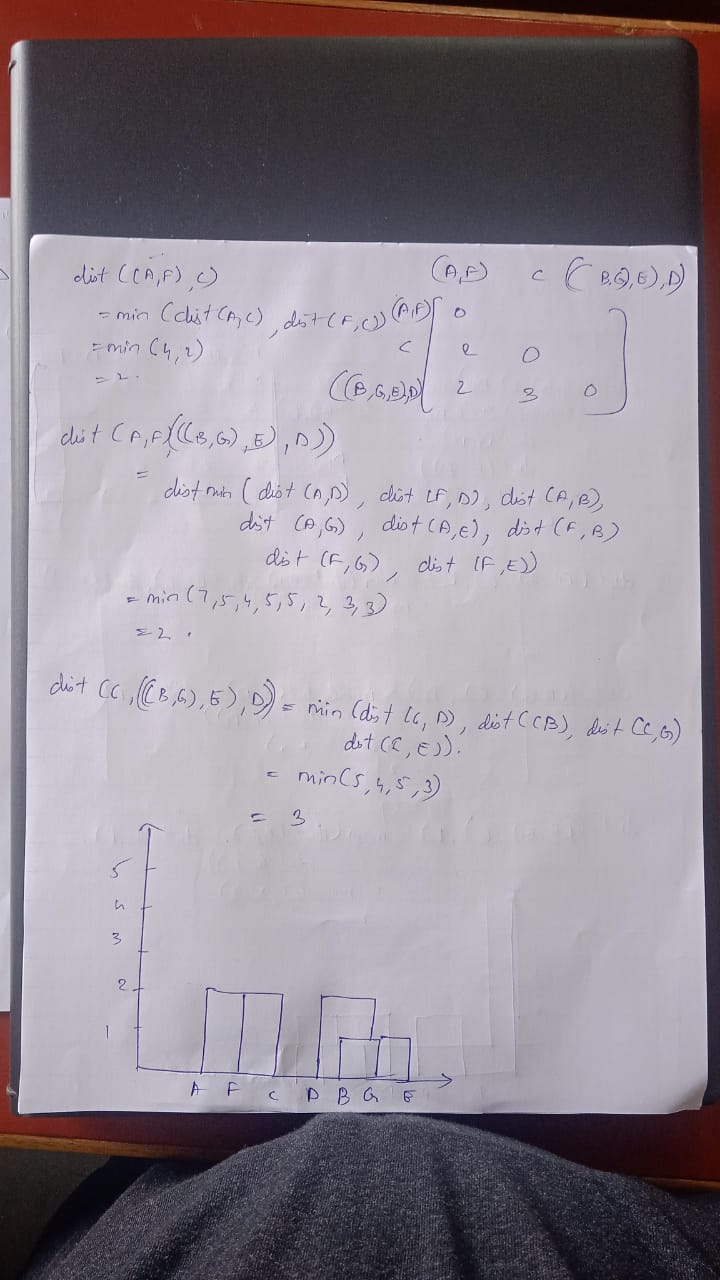
F(2,3)

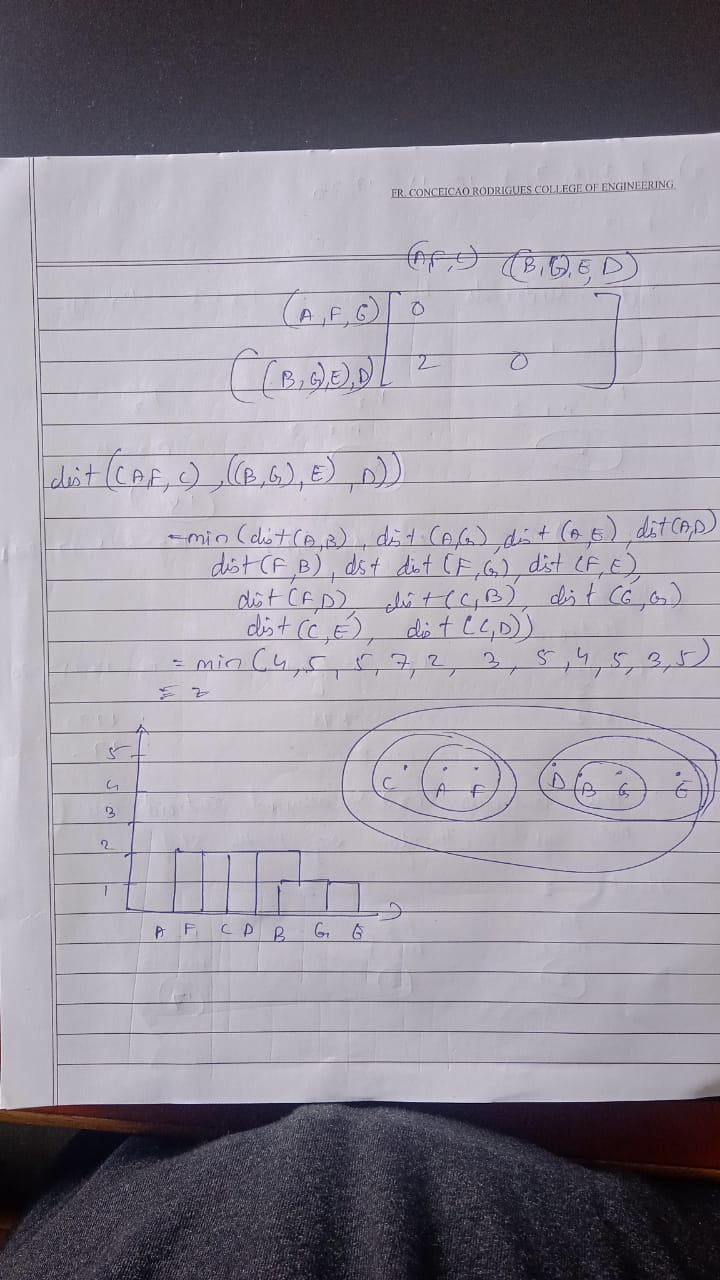
G(3,5)







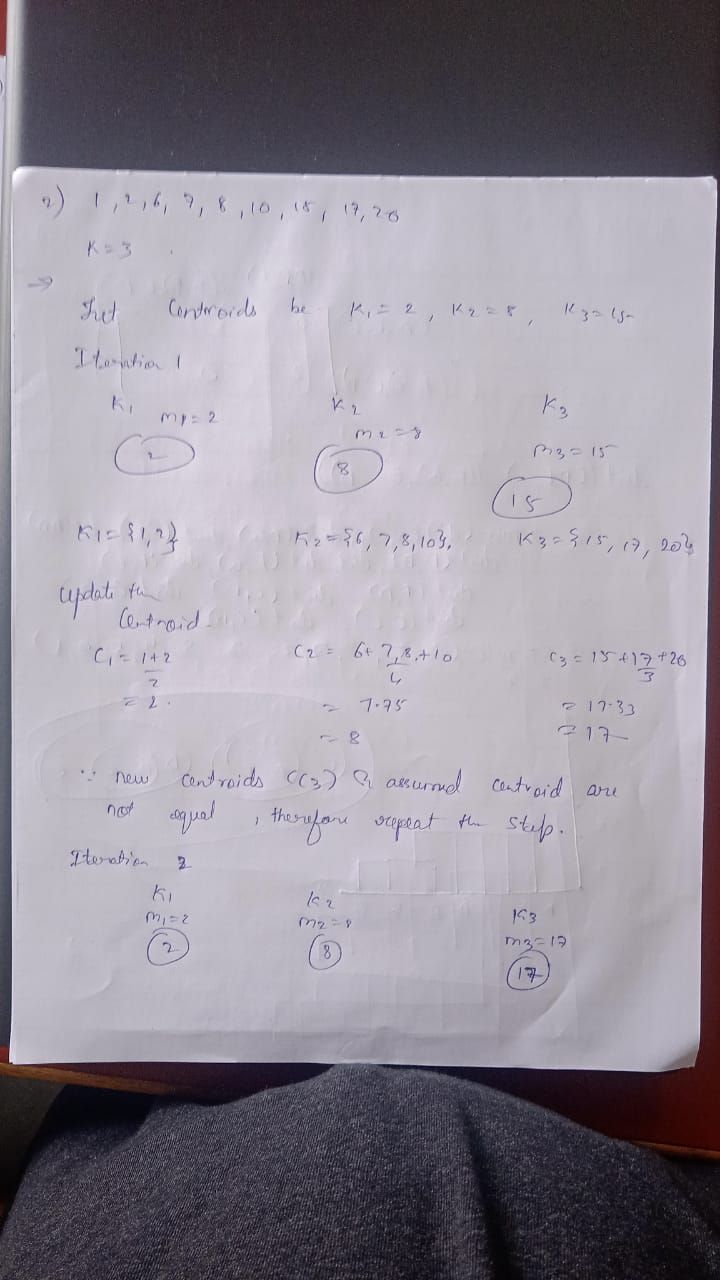


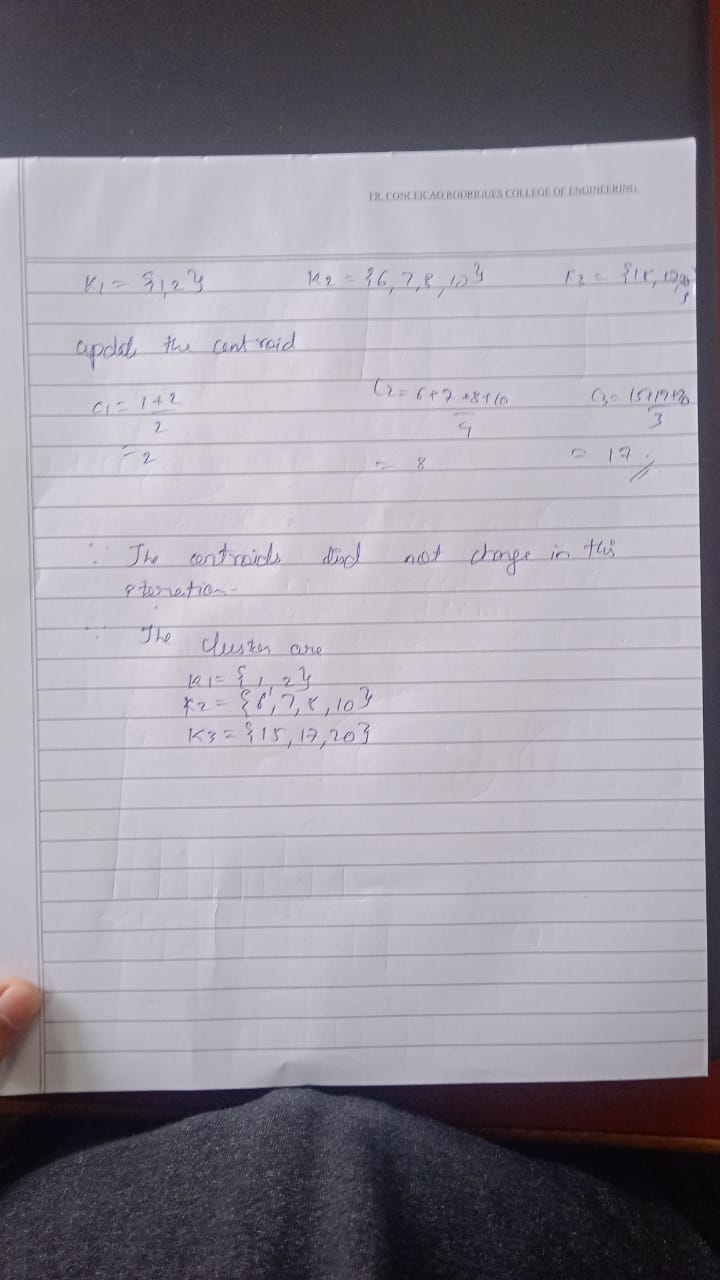


1. Apply k-mean clustering technique on following dataset.

1,2,6,7,8,10,15,17,20

K=3

****

****

**Conclusion:**

**We are able to understand different clustering algorithm and am able to implement the algorithm using WEKA tool.**