	Hope Foundation's **Finolex Academy of Management and Technology, Ratnagiri** **Department of Information Technology**		
Subject name	**Business Intelligence Lab**		**Subject Code: ITL602**
Class	**TE IT**	**Semester – VI (CBCGS)**	**Academic year: 2018-19 (FH 2019)**
Name of Student			**QUIZ Score :**
Roll No		**Assignment/Experiment No:**	**03**
Title:	**To use WEKA to implement Clustering Algorithms**		

1. Course objectives applicable: LO3

2. Course outcomes applicable: LO3

3. Learning Objectives:

1. To learn clustering techniques
2. To understand unsupervised learning.
3. To learn to run clustering algorithms in WEKA

4. Practical applications of the assignment/experiment: Cluster analysis of data collection

5. Prerequisites:

1. Types of Clustering algorithms
2. Unsupervised learning concepts

6. Hardware Requirements:

1. PC with minimum 2 GB RAM

7. Software Requirements:

1. Windows 8.1 or higher
2. WEKA 3.8 or higher

8. Viva Questions (if any): (Online Quiz will be taken separately batch-wise)

1. What is clustering?
2. What is unsupervised learning?
3. What are the different types of clustering?
4. How clustering is different than classification?

9. Experiment/Assignment Evaluation:

Sr. No.	Parameters	Marks obtained	Out of
1	Technical Understanding (Assessment may be done based on Q & A **or** any other relevant method.)		6
2	Neatness/presentation		2
3	Punctuality		2

Date of performance (DOP)		**Total marks obtained**		**10**
Date of checking (DOC)		**Signature of teacher**		

10. Theory: **<<handwritten work>>**

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.

In clustering -
1. Intra cluster distance between objects is minimum.
2. Inter cluster distance between objects is maximum.

The types of clustering in different area are.

1. Connectivity-based clustering (hierarchical clustering)
2. Centroid-based clustering (k-means, k-medoids)
3. Distribution-based clustering (probability distributions, Gaussian distribution)
4. Density-based clustering (DBSCAN, OPTICS)

11. Performance Steps: **<<Handwritten>>**

1. Start WEKA
2. Open the data set (ARFF file)
3. Select the clustering technique (k-means, agglomerative, divisive) and run the trial
4. Record the reading and find precision and Recall
5. Identify a subset of instances having improved clustering performance.

12. Results:

<< Add the hard-copy of output screen shots >>

13. Learning Outcomes Achieved

1. Understanding of unlabeled data and unsupervised learning.
2. Understanding the clustering techniques and their outcomes.
3. Understanding different clustering types and their application area.

14. Conclusion:

1. **Applications of the studied technique in industry:** To perform data analysis.
2. **Engineering Relevance:** To build data analysis software using clustering.
3. **Skills Developed:** Understating of clustering techniques and their applications

15. References:

[1] Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3rd Edition.

[2] P. N. Tan, M. Steinbach, Vipin Kumar, "Introduction to Data Mining", Pearson Education.

[3] Michael Berry and Gordon Linoff, "Data Mining Techniques", 2nd Edition Wiley Publications

[4] https://en.wikipedia.org/wiki/Cluster_analysis

[5] https://en.wikipedia.org/wiki/Unsupervised_learning