COURSE HANDOUT

B. Tech (CSE) – 8th Semester(FSI)

(Applicable for the Batches Admitted from 2014-15, Non – FSI & FSI Model)

Course Title	: **Data Engineering Lab**	Dated	: 20-11-2017
Course Code	: CSE 4226	Academic Year	: 2017-18
Course Structure	: 3-1-0-4		
Course Coordinator	**:** Dr. S.S.Gantayat		
Instructor(s)	**:** Dr. S.S.Gantayat		

Course Objective:
The course content enables students to:

1. Conceptualize the Data Mining Problem

2. Perform Preprocess Data

3. Analyze and Visualize the Data with Data Mining Techniques

4. Perform Predictive Modeling

5. Generate Association Rules for Business Data

Course Outcome:
At the end of the course students will be able to:

1. Implement the Algorithms to solve Data Mining problem using WEKA tool

2. Identify an appropriate method to apply in a given situation

3. Communicate results in terms relevant to Science, Business etc.

4. Apply different classification and clustering techniques to characterize subgroups.

Lab Manuals:
Text Books:

1. Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Pearson Education, 2002.

2. Jiawei Han & Micheline Kamber, Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann, India, 2010.

Reference Books:

1. Data Mining: Introductory and Advanced Topics, Margaret H. Dunham, Pearson Education. 2002.
2. Data Mining:Practical Machine Learning Tools and Techniques, Ian H. Witten, Eibe Frank, Mark A. Hall, 3rd Edition, Morgan Kaufmann Publishers, 2011.
3. Arun K. Pujari, Data Mining Techniques, 3rd Edition, University Press. 2013.

Other References:

Weka Examples: https://svn.scms.waikato.ac.nz/svn/weka/branches/stable-3-6/wekaexamples/
Weka Manual : http://statweb.stanford.edu/~lpekelis/13_datafest_cart/WekaManual-3-7-8.pdf
Weka Software: http://www.cs.waikato.ac.nz/ml/weka/downloading.html

SYLLABUS: (List of Experiments)

Softwares to Implement: **WEKA/ Clementine/ IBM SPSS**

Experiment 1.

Introduction to Graphical User Interface (GUI) of WEKA.

Experiment 2.

Perform Data Pre-processing on a sample data set.

Experiment 3.

Introduction to IBM SPSS Modeler and nodes palette.

Experiment 4.

Preparing the data for analysis using data audit node.

Experiment 5.

Automated data preparation using data audited node.

Experiment 6.

Perform Association Analysis to derive the association rules Algorithm.

Experiment 7.

Implement the Classification using Decision Tree Induction Algorithm.

Experiment 8.

Implement the Classification using Regression.

Experiment 9.

Implement the Bayesian Classification Algorithm.

Experiment 10.

Classify Telecommunications Churn by using Binomial Regression Algorithm.

Experiment 11.

Implement Market Basket Analysis using Rule Induction/C5.0 Algorithm.

Experiment 12.

Predicting Loan Defaulters using Bayesian Networks Algorithm

Experiment13.

Implement K-means clustering Algorithm.

Experiment 14.

Implement Hierarchical clustering Algorithm.

Course Plan:

Experiment No.	Learning Objectives	Topic(s) to be Covered
1	To understand the different Graphics User Interfaces of the WEKA Software package for implementation.	Introduction to Graphical User Interface (GUI) of WEKA
2	To perform data pre-processing for some sample dataset	Perform Data Pre-processing on sample data set
3	To understand the different Graphics User Interfaces of the IBM SPSS Modeler and nodes palette for implementation.	Introduction to IBM SPSS Modeler and nodes palette
4	To prepare data for analysis using data audit nodes.	Preparing the data for analysis using data audit node.
5	Implementing automated data preparation using data audited node.	Automated data preparation using data audited node.
6	To perform Association Analysis to derive the association rules algorithm for a given data set	Perform Association Analysis to derive the association rules.
7	To implement the Classification using Decision Tree Induction algorithm.	Implementing the Classification using Decision Tree Induction algorithm.
8	To implement the Classification using Regression algorithm	Implementing the Classification using Regression algorithm.
9	To implementing the Bayesian Classification algorithm for given sample data.	Implementing the Bayesian Classification Algorithm.
10	To classifying telecommunications churn by using Binomial Regression algorithm	Classifying telecommunications churn by using Binomial Regression algorithm.
11	To analyze Market Basket Analysis using Rule Induction/C5.0 Algorithm	Analyzing Market Basket Analysis using Rule Induction/C5.0 Algorithm
12	To predict Loan Defaulters using Bayesian Networks Algorithm from a bank data.	Predicting Loan Defaulters using Bayesian Networks Algorithm.
13	To implement the K-means clustering algorithm for a given data set.	Implement K-means clustering Algorithm
14	To implement the Hierarchical clustering algorithm for a given data set.	Implement Hierarchical clustering Algorithm.

Evaluation Scheme:

Component	Particulars	Marks	Date & Time
Lab Regularity	No. of Experiments completed and recorded	15	Every week during the semester
Internal Examination	150 Minutes	10	26-03-2018 to 31-03-2018
External Examination	180 Minutes	50	02-04-2018 to 07-04-2018
Total		**75**	

Venue :   Programming Lab / Networks Lab
Notices: CSE Main Notice Board

S. S. Gantayat
Signature of the Instructors

S. S. Gantayat
Signature of the Course-Coordinator