Academic Year: **2018-19** Subject: **Data Mining Lab (16ITC29)** Semester: **VI (H2)**

S.No	Week	To include in the record
1.	10.12.2018 – 12.12.2018	Introduction to data mining using Weka and R-Tool i. Write short notes on the below options available WEKA GUI • Explorer • Experimenter • Knowledge Flow • Workbench • Simple CLI ii. Preprocessing • Attribute Selection • Handle Missing Values (labour data set of weka/ or any dataset from keel repository) • Nominal to binary • Discretisation • Normalisation • Standardisation • Data Visualisation (Sequence of options that you need to choose to apply a particular filter, description about each filter and the different parameters available with each filter) iii. About .arff files and its format iv. Description about any two data sets v. http://weka.sourceforge.net/doc.stable/ - write class hierarchy and description about any two filters that you have tried R-Tool vi. Commands to read.csv file vii. Commands to write a .csv file viii. Data structures: data frame, list, matrices, arrays, slicing of lists ix. List of packages available in R-Tool which supports data mining functionalities *Supporting Screenshots are must

2.	17.12.2018-19.12.2018	i. Weka.filters.unsupervised.attribute.InterquartileRange ii. Outlier detection and elimination using weka iii. Data Exploration and Visualization in R (Help from: http://www.rdatamining.com/examples/exploration , Rtool ppt/Gardener text book) • Check the dimensionality of the chosen dataset • Variable names or column names • Structure • Attributes • Get the first 5 rows • Get the last 6 rows • Get second attributes of the first 10 rows • Distribution of every dimension • Frequency of each class type • Pie chart • Variance of a numeric attribute • Covariance of two attributes • Correlation of two dimensions • Histogram of an attribute • Density • Scatter plot • Pair Plot • Box-Whisker Plots • Line charts for both numeric and categorical dimensions • Cleveland Dot Charts • Bar Charts (Note your observations, Comment on the data distribution, try plotting commands for different kinds of dimensions, try different plotting function options: symbols, size of plotting symbol, legends, x,y-axis labels, titles of graphs, etc)

		iv. Start working on a project in Python (Can be using opencv, tensorflow, etc…) preferably individual projects not in teams
3.	24.12.2018 – 26.12.2018	i. Generate Association Rules using Apriori and FP Growth algorithms in WEKA (http://www.rdatamining.com/examples/association-rules) ii. Generate Association Rules using Apriori, ECLAT and FP Growth* algorithms Using R-Tool iii. Convert .csv to .arff and vice versa using R-Tool, WEKA and JAVA API iv. Start working on the project v. Be ready for viva vi. Get your record with first two week tasks and respective solutions (Try different arguments in R-Tool) * Document mailed
4.	31.12.2018 – 02.01.2019	i. Complete pending programs and work on projects ii. Make sure that you include comments for R-programs for every single line in the program iii. Try different arguments iv. Write complete syntax with purpose of each argument for the commands in R v. Complete record vi. Work on your project
5.	07.01.2019 - 09.01.2019	i. Figure out the datasets and decide the task(s) ii. Write complete syntax with purpose of each argument for the commands in R iii. Complete record
6.		Classification i. Decision Tree ii. Naïve Bayes iii. Bagging iv. AdaBoost v. Random forest vi. vi. K-NN
7.		Clustering: (In WEKA as well as R-Tool) i. K-Means ii. Hierarchical

		iii. DBSCAN
8.		Fashion MNIST - Tensorflow
9.		Build Linear Regression model – Using R-Tool
10.		Text Mining with R: Twitter Data Analysis Word cloud Sentiment analysis
11.		Time Series Analysis and Mining Forecasting Clustering Classification

Resources

1. Association Rule Mining
    a. https://rdrr.io/cran/rCBA/man/fpgrowth.html
    b. http://r-statistics.co/Association-Mining-With-R.html
    c. http://www.salemmarafi.com/code/market-basket-analysis-with-r/
    d. http://rstatistics.net/association-mining-with-r/
    e. http://www.borgelt.net//fpgrowth.html
2. Data Exploration and Visualisation in R
    a. https://r4ds.had.co.nz/exploratory-data-analysis.html
3. Text Mining
    a. https://rdrr.io/rforge/tm/man/content_transformer.html
    (S.No#discussed in Class)
    b. http://dataaspirant.com/2018/03/22/twitter-sentiment-analysis-using-r/
    (Sentiment Analysis)
    c. http://www.rdatamining.com/docs/text-mining-with-r
    (According to syllabus)
    d. #http://www.sthda.com/english/wiki/text-mining-and-word-cloud-fundamentals-in-r-5-simple-steps-you-should-know

(Word cloud/Text visualisation)

4. Time series Analysis
    a. http://www.rdatamining.com/docs/time-series-analysis-and-mining-with-r
    b. https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/
    c. https://www.analyticsvidhya.com/blog/tag/time-series-analysis/
    d. https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/
    e. https://www.quora.com/How-do-I-learn-about-time-series-analysis-2
    f. https://towardsdatascience.com/end-to-end-time-series-analysis-and-modelling-8c34f09a3014
    g. https://towardsdatascience.com/analyzing-time-series-data-in-pandas-be3887fdd621
    h. https://en.wikipedia.org/wiki/Dynamic_time_warping
5.