## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA KAKINADA – 533 003, Andhra Pradesh, India

#### **DEPARTMENT OF INFORMATION TECHNOLOGY**

III Year – I Semester		L	Τ	Р	C	
III Year – I Semester		0	0	3	1.5	
DATA MINING TECHNIQUES WITH R LAB						

#### **Course Objectives:**

- To understand the mathematical basics quickly and covers each and every condition of data mining in order to prepare for real-world problems.
- The various classes of algorithms will be covered to give a foundation to further apply knowledge to dive deeper into the different flavors of algorithms.
- Students should aware of packages and libraries of R and also familiar with functions used in R for visualization.
- To enable students to use R to conduct analytics on large real life datasets.
- To familiarize students with how various statistics like mean median etc. can be collected for data exploration in R.

#### Course Outcomes(COs): At the end of the course, student will be able to

- Extend the functionality of R by using add-on packages
- Extract data from files and other sources and perform various data manipulation tasks on them.
- Code statistical functions in R
- Use R Graphics and Tables to visualize results of various statistical operations on data
- Apply the knowledge of R gained to data Analytics for real life applications

#### List of Experiments:

Implement all basic R commands.

Interact data through .csv files (Import from and export to csv files)

Get and Clean data using swirl excercises. (Use 'swirl' package, library and install that topic from-swirl).

Visualize all Statistical measures (Mean, Mode, Median, Range, Inter Quartile Range etc., using Histograms, Boxplots and Scatter Plots).

Create a data frame with the following structure.

EMP ID	EMP NAME	SALARY	START DATE
1	Satish	5000	01-11-2013
2	Vani	7500	05-06-2011
3	Ramesh	10000	21-09-1999
4	Praveen	9500	13-09-2005
5	Pallavi	4500	23-10-2000

Extract two column names using column name

Extract the first two rows and then all colun

Extract 3<sup>rd</sup> and 5<sup>th</sup> row with 2<sup>nd</sup> and 4<sup>th</sup> column

Write R Program using 'apply' group of function in the program alization function on each of the numeric variables/columns of iris dataset to transform them into

i. 0 to 1 range with min-max normalization.

ii. a value around 0 with z-score normalization.

Create a data frame with 10 observations and 3 variables and add new rows and columns to itusing 'rbind' and 'cbind' function.

Write R program to implement linear and multiple regression on intears' dataset to estimate the value of 'mpg' variable, with best R<sup>2</sup> and plot the original values in 'green' and predicted values in 'red'.



# Authentity Technology

## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA KAKINADA – 533 003, Andhra Pradesh, India

### DEPARTMENT OF INFORMATION TECHNOLOGY

Write R Programs to implement k-means clustering
Write R Programs to implement k medoids channeling
Write R Programs to implement density based clustering on iris dataset
Write a R Program to implement decision nees using 'readingSkills' dataset
Implement decision trees using 'iris' dataset using package party and 'rpart'
Use a Corpus() function to create a data corpus then Build a ferm Matrix and Reveal word frequencies.
Write a R Program to implement decision trees using 'readingSkills' dataset   Implement decision trees using 'iris' dataset using package party and 'rpart'   Use a Corpus() function to create a data corpus then Build a term Matrix and Reveal word frequencies.

#### **Reference Books:**

- 1. <u>www.tutorialspoint.com/r</u>
- 2. <u>www.r-tutor.com</u>
- 3. R and Data Mining: Examples and Case Studies, 1<sup>st</sup> ed, Yanchang Zhao, Sprnger, 2012.
- 4. https://towardsdatascience.com/