**MS RAMAIAH INSTITUTE OF TECHNOLOGY**

(AUTONOMOUS INSTITUTE, AFFILIATED TO VTU)

BANGALORE - 560 054

**SEMESTER END EXAMINATIONS - JANUARY 2016****M.Tech:- Computer Science and**

Course &amp; Branch :

**Engg. /**Semester : **III****Computer Network Engg.**Subject : **Big Data and Data Science**Max. Marks : **100**Subject Code : **MCSEE12/MCNEE12**Duration : **3 Hrs****Instructions to the Candidates:**

- Answer one full question from each unit.

**UNIT - I**

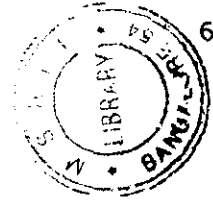
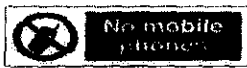
1. a) Describe the traditional data analytical architecture with a neat diagram. In which phase would the team expect to invest most of the project time? Why? (10)
- b) Describe the tools that are used in data preparation and model planning phase in the data analytics life cycle. (10)
- 2 a) Explain the major groups that are evolving in the emerging Big data ecosystems and key roles in an analytics project. (10)
- b) Explain the overview of data analytics life cycle by highlighting the different users who are responsible in the life cycle. (10)

**UNIT - II**

3. a) Describe how clustering can be carried out using k-means clustering method. (10)
- b) Describe the use cases of Association rules and contrast the differences between k-means clustering and Apriori algorithm. (10)
4. a) Illustrate Apriori algorithm with the different measures used in determining the appropriateness of the candidate rules. (10)
- b) Describe the use cases of k-means clustering along with its disadvantages. (10)

**UNIT - III**

5. a) Describe how linear regression can be carried out using Ordinary least squares method. (10)
- b) Explain how you detect significant splits using a decision tree. (10)



6. a) Describe ID3 and C4.5 decision tree algorithm. What are the differences? (10)
- b) Describe naïve bayes classification using Bayes theorem. (10)

**UNIT - IV**

7. a) Explain how you carry out time series analysis using Box - Jenkins methodology along with ARIMA model. (10)
- b) Explain how you calculate term frequency for text analysis with an example. (10)
8. a) Explain how time series can be carried out using Auto regressive models. (10)
- b) Describe POS tagging, Lemmatization, and Stemming with an example along with different sources of data available for text analysis. (10)

**UNIT - V**

9. a) Describe the architecture of HDFS showing how files are stored in HDFS. (10)
- b) Describe the usage of window functions that can be used for advanced SQL analytics. (10)
10. a) Explain the structure of map reduce program with an example. (10)
- b) Describe EWMA advanced SQL time series analytical method with an example. (10)

\*\*\*\*\*