

(2 ½ Hours)

[Total Marks: 75]

N.B. : 1) All questions are **compulsory**.2) **Figures** to the right **indicate marks**.3) **Illustrations**, in-depth **answers** and **diagrams** will be **appreciated**.4) Mixing of **sub-questions** is **not allowed**.**Q.1 Attempt All(Each of 5marks)****(15M)****(a) Choose the correct alternative.**

- i. Ordinal data is a type of _____.
 - a) Measurement
 - b) Catagorical
 - c) Discrete
 - d) Continuous
- ii. _____ measures asymmetry about the mean of the probability distribution of a random variable.
 - a) skewness
 - b) covariance
 - c) variance
 - d) Kurtosis
- iii. _____ shows all individual data points.
 - a) Box-plot
 - b) scatter plot
 - c) line plot
 - d) pie chart
- iv. In _____, we start with all the features and removes the least significant feature at each iteration.
 - a) Forward elimination
 - b) Backward elimination
 - c) Recursive Feature elimination
 - d) None of the above
- v. XPath specification has _____ types of nodes
 - a) Four
 - b) Five
 - c) Six
 - d) Seven

(b) Fill in the blanks. Use following pool to answer question.

Pool (XPath, Mode, sample, Tidy Data, Principal Components Analysis)

- i. _____ is the subset of the population.
- ii. Five number summary does not include _____ value.
- iii. _____ refers to the process by which principal components are computed.
- iv. _____ is a query language that is used for traversing through an XML document.
- v. _____ data has each variable saved in its own column.

(c) Answer in 1-2 sentences.

- i. Define Mode.
- ii. What is Data Aggregation?
- iii. What is Margin in SVM?
- iv. What is feature extraction?
- v. Define BIC.

Q.2 Attempt the following(Any Three)

(15M)

- Explain data along with its types.
- Describe the types of observational methods used in data collection.
- What is EDA? Explain any two types of visualization.
- Compare structured and unstructured data.
- What is kurtosis? Explain its types.
- What is data wrangling? Explain with any one package.

Q. 3 Attempt the following (Any THREE) (Each of 5Marks)

(15M)

- What is NoSQL? Briefly explain its types.
- What is JSON? How to read JSON file in R with an example?
- Write a note on HBase.
- What is collection in MongoDB? Give an example to create collection in MongoDB.
- What are various ways to perform Web Crawling?
- Write note on XPath.

Q. 4 Attempt the following (Any THREE) (Each of 5Marks)

(15)

- Write short note on Time Series Analysis.
- Explain Multiple Linear Regression.
- What is Decision tree? What are its advantages?
- Write note on Bias/Variance Tradeoff?
- Explain hierarchical clustering.
- Write a short note on Ensemble Methods.

Q. 5 Attempt the following (Any THREE) (Each of 5Marks)

(15)

- Distinguish between Structured and Unstructured Data.
- Explain Map Reduce Architecture.
- Write note on Support vector Machines.
- Explain Bayesian Information Criterion.
- What is MongoDB? State its features.
