YADAVA COLLEGE(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE

BIG DATA ANALYTICS

CLASS: **SEMESTER**:

UNIT I

1. Which of the following is one of the 3Vs of Big Data?
   a) Volume b) Accuracy
   c) Validity d) Reliability
2. Big Data evolved primarily due to:
   a) Increase in structured data **b) Growth of unstructured data**
   c) Decline in storage costs d) Reduction in computing power
3. Which component stores large-scale data in Big Data systems?
   a) MongoDB b) SQL Server
   c) Oracle DB **d) Hadoop Distributed File System**
4. Which type of analytics predicts future outcomes?
   a) Descriptive analytics b) Diagnostic analytics
   c) Prescriptive analytics **d) Predictive analytics**
5. A major challenge in Big Data is:
   a) Lack of data **b) Data privacy and security**
   c) Limited applications d) Small datasets
6. Which skill is essential for Big Data professionals?
   a) Cooking **b) Data visualization**
   c) Gardening d) Sports management
7. Which is a common Big Data application?
   a) Image compression b) Database indexing
   c) Recommendation systems d) File encryption
8. Which type of analytics explains why something happened?
   a) Descriptive **b) Diagnostic**
   c) Predictive d) Prescriptive
9. Which is NOT one of the 3Vs of Big Data?
   a) Volume b) Variety
   c) Velocity **d) Validity**

10. Which industry uses Big Data for personalized marketing?

a) Retail b) Agriculture

c) Mining d) Construction

UNIT II

11. Classification algorithms are used for:

a) Predicting continuous values **b) Grouping data into categories**

c) Data visualization d) Data storage

12. Regression techniques are used to:

a) Predict categorical outcomes **b) Predict numerical values**

c) Classify text d) Cluster data

13. Text analytics is useful for:

a) Image recognition b) Hardware Design

c) Sentiment analysis d) Network routing

14. Real-time systems are characterized by:

a) Delayed response **b) Immediate response within deadlines**

c) No response d) Random response

15. Which is a type of real-time system?

a) Hard real-time b) Batch processing

c) Cloud computing d) Data warehousing

16. Hadoop is mainly used for:

a) Real-time processing **b) Batch processing**

c) Image editing d) Video streaming

17. Real-time analytics is crucial for:

a) Stock market monitoring b) Recipe suggestions

c) Gardening tips d) Sports commentary

18. In-database analytics means:

a) Analytics outside the database **b) Analytics within the database engine**

c) Ignoring analytics d) Using spreadsheets

19. Real-time system architecture includes:

a) Input, processing, output b) Only input

c) Only output d) Input, Processing

20. Healthcare analytics may include:

a) Patient risk prediction b) Movie recommendations

c) Weather forecasting d) Sports scores

UNIT III

21. Virtualization in Big Data helps in:

a) Enhance gardening b) Cooking faster

c) Improving sports **d) Reducing hardware dependency**

22. Big Data stack includes:

a) Only visualization b) Only storage

c) Only analytics **d) Data storage, processing, analytics**

23. High-dimensional data refers to:

a) Few attributes **b) Many attributes**

c) No attributes d) Random data

24. Dimensionality reduction is used to:

a) Increase features **b) Reduce features while retaining info**

c) Eliminate all data d) Duplicate data

25. PCA is used for:

a) Dimensionality reduction b) Data encryption

c) Data storage d) Data visualization only

26. Curse of dimensionality occurs when:

a) Too few features **b) Too many features**

c) Missing data d) Encrypted data

27. Which is NOT a dimensionality reduction technique?

a) PCA b) LDA

c) Decision Trees d) t-SNE

28. Big Data hardware includes:

a) Distributed servers b) Printers

c) Scanners d) Monitors

29. Virtualization allows:

a) Multiple OS on one machine b) Only one OS

c) No OS d) Random OS

30. Dimensionality reduction improves:

a) Model accuracy and efficiency b) only accuracy

c) only speed d) only efficiency

UNIT IV

31. Which is NOT a basic data type in R?

a) Numeric b) Character

c) Boolean **d) Cooking**

32. Vectors in R are :

a) Homogeneous b) Heterogeneous

c) Random d) Empty

33. Lists in R can contain:

a) Only numbers b) Only characters

c) Mixed data types d) Only logical values

34. Factors in R are used for:

a) Numerical computation **b) Categorical data**

c) Image processing d) Encryption

35. Arrays in R are:

a) Multi-dimensional b) One-dimensional

c) Empty d) Random

36. Control statements in R include:

a) if, else, switch b) loop, break, return

c) Both a and b d) start, stop

37. Data frames in R are similar to:

a) Tables b) Images

c) Videos d) Sounds

38. Which operator is used for assignment in R?

a) <- b) =

c) -> **d) Both a and b**

39. Loops in R include:

a) for, while, repeat b) if, else

c) switch, case d) return

40. Which statement exits a loop in R?

a) break b) continue

c) return d) stop

UNIT V

41. Function to import data in R:

a) read.csv() b) write.csv()

c) import() d) load()

42. Exporting data to CSV in R uses:

a) read.csv() b) save()

c) export.csv() **d) write.csv()**

43. Handling missing values in R:

a) na.delete() b) na.remove()

c) na.omit() d) na.ignore()

44. Function to calculate mean in R:

a) mean() b) avg()

c) calculate() d) sum()

45. R graphics are generated using:

a) draw() b) graph()

c) chart() **d) plot()**

46. ggplot2 is used for:

a) Data visualization b) Data storage

c) Data encryption d) Data import

47. Function for histogram in R:

a) plot() b) bar()

c) hist() d) graph()

48. Missing values in R are represented as:

a) -1 b) NULL

c) 0 **d) NA**

49. Data visualization helps in:

a) Understanding patterns b) storing patterns

c) deleting patterns d) writing patterns

50. Function to summarize data in R:

a) summary() b) describe()

c) overview() d) details