R6676	Sub. Code
	7BD2C1

P.G. DIPLOMA EXAMINATION, APRIL – 2022

Second Semester

Big Data Analytics

ADVANCED BIG DATA TECHNOLOGIES

(CBCS – 2018 onwards)

Time : 3 Hours Maximum : 75 Marks

Part A $(10 \times 2 = 20)$

Answer **all** questions.

1. Define veracity of big data.

2. Mention any four languages used for writing queries for processing big data.

3. Write any two differences between HDFS, GFS.

4. Define data consistency for bid data.

5. What is the storage model used by Amazons Dynamo?

6. List any four problems found in big data visualization.

7. Define scalable algorithm.

8. Define centrality and what is the use of centrality algorithms.

9. What are the three execution models of Hadoop?

10. What is the purpose of input split?

Part B (5 × 5 = 25)

Answer **all** questions, choosing either (a) or (b).

11. (a) Write short notes on how big data is stored and processed.

Or

(b) What are the challenges in real time big data analytics?

12. (a) What are the operations carried out by GFS architecture client?

Or

(b) Write short notes on limitations and advantages of Hadoop.

13. (a) Explain briefly about document storage model.

Or

(b) Explain briefly about structured vs unstructured data.

14. (a) Write short notes on community detection.

Or

(b) Write briefly about the frequent item detecting algorithm and analyse it.

15. (a) Write notes on server farms and how distributed data processing is carried across server farm.

Or

(b) Write notes on building blocks of Hadoop mapreduce.

Part C

$(3 \times 10 = 30)$

Answer any **three** questions.

16. Explain big data stack architecture.

17. Explain Hadoop architecture.

18. Explain compliance challenges associated with big data.

19. Explain the features of Pig Latin Hadoop language.

20. Explain the steps in executing Hadoop mapreduce job.

———————

R6677	Sub. Code
	7BD2C2

P.G. DIPLOMA IN BIG DATA ANALYTICS EXAMINATION, APRIL – 2022

Second Semester

ALGORITHMS FOR ADVANCED ANALYTICS

(CBCS – 2018 onwards)

Time : 3 Hours Maximum : 75 Marks

Part A $(10 \times 2 = 20)$

Answer **all** the questions.

1. List the two-step process of Data classification.

2. What is back propagation?

3. Define Decision Tree.

4. What is information gain?

5. What is document collection in text mining?

6. Define Clustering.

7. What are support vector machines?

8. Expand FAM.

9. Define linear discrimination.

10. What is Feed-forward network mappings?

Part B (5 × 5 = 25)

Answer **all** questions, choosing either (a) or (b).

11. (a) Compare and Contrast about classification and prediction methods.

Or

(b) Explain about rough set approach for classification.

12. (a) Explain the basic algorithm for inducing a decision tree from training tuples.

Or

(b) Write short note on attribute selection measures.

13. (a) Explain any five core text mining operations.

Or

(b) Discuss the three common text categorization in detail.

14. (a) Explain the basics of Neural Networks in detail.

Or

(b) Write short note on appropriate rules for the functions to be modelled.

15. (a) Explain the sum-of-squares error function.

Or

(b) Discuss the binary inputs and continuous inputs.

Part C (3 × 10 = 30)

Answer any **three** questions.

16. Explain how the Naive Bayesian classifier works in detail.

17. Describe the various classification of decision tree induction.

18. Explain the document representation of text categorization.

19. Discuss the advantages and disadvantages of Neural Networks.

20. Describe the perceptron convergence theorem in detail.

———————

R6678	Sub. Code 7BD2C3

P.G. DIPLOMA EXAMINATION, APRIL – 2022

Second Semester

Big Data Analytics

BIG DATA SECURITY AND RISK ANALYSIS

(CBCS – 2018 onwards)

Time : 3 Hours Maximum : 75 Marks

Part A (10 × 2 = 20)

Answer **all** questions.

1. What is meant by authorization and authentication?
2. Define network security.
3. What is Hive?
4. Define Kerberos.
5. What is meant by Risk?
6. Define Risk Management.
7. List any two usages of big data in Bank domain.
8. Draw Root cause of the crisis.
9. What is meant by Actuary services?
10. What is Health value and cost?

Part B (5 × 5 = 25)

Answer **all** questions, choosing either (a) or (b).

11. (a) Write a short note on Ethical guidelines.

Or

(b) List down the steps involved in securing big data.

12. (a) How to configure kerberos for Hadoop ecosystem components?

Or

(b) Discuss on Pig.

13. (a) Elaborate Risk Mitigation.

Or

(b) Write a short note on Risk analytics.

14. (a) How can big data helps in financial crisis?

Or

(b) Write a note on credit Risk analytics.

15. (a) Write a short note on customer insights.

Or

(b) How to handle risk management in healthcare domain?

Part C (3 × 10 = 30)

Answer any **three** questions.

16. Explain in detail about overview of security (CIA).

17. Describe in detail about Hadoop kerberos security.

18. Explain about Risk assessment.

19. Explain in detail on handling fraud risk in the bank domain.

20. Discuss the usage of big data in healthcare sector.

———————

R6679	Sub. Code
	7BD2G3

P.G. DIPLOMA EXAMINATION, APRIL – 2022

Second Semester

Diploma in Big Data Analytics

WEB INTELLIGENCE AND SOCIAL NETWORK ANALYSIS

(CBCS – 2018 onwards)

Time : 3 Hours Maximum : 75 Marks

Part A (10 × 2 = 20)

Answer **all** questions.

1. List any two example for intelligent web applications

2. What is the key concept of network analysis?

3. Define sentiment detection.

4. What is vertical search engine?

5. What is a structural hole?

6. Define Clique.

7. What is homophily in social relations?

8. What is positive relationship?

9. How is Web community extracted?

10. Define growth rate.

Part B (5 × 5 = 25)

Answer **all** questions, choosing either (a) or (b).

11. (a) Briefly explain the development of Social Network Analysis.

Or

(b) Explain the statistical properties of social network analysis.

12. (a) Write the properties of Naïve Bayes.

Or

(b) Explain about Clustering issues in very large data sets.

13. (a) Write the Varieties of Social Networks.

Or

(b) Write about Network datasets.

14. (a) Explain about negative relationships.

Or

(b) Write about structural balance.

15. (a) Explain about Analyzing Communities and Evolutions in Dynamic Network.

Or

(b) Elaborate Complete Bipartite Graphs.

Part C (3 × 10 = 30)

Answer any **three** questions.

16. What are the limitations of current Web? Explain the development of semantic Web and the emergence of Social Web.

17. Compare various feature selection methods.

18. Explain about Clustering of Social-Network Graphs.

19. Discuss about spatial model of segregation.

20. Explain about the Girvan-Newman Algorithm.

———————