19BDH301

Reg No : .....

CHOICE BASED CREDIT SYSTEM
M.Sc. THIRD SEMESTER DEGREE EXAMINATION FEBRUARY 2022
BIG DATA ANALYTICS
Big Data Analytics using Spark

Duration:3 Hrs **Max Marks:70**

PART - A

1. Answer any EIGHT of the following: **(2×8= 16)**

a) Discuss the two parameters that are passed to initialize a SparkContext.

b) Enumerate the common supported file formats in Spark.

c) Which is the class used for the primary configuration mechanism in Spark?

d) What do you understand by the term SchemaRDD ?

e) Summarize the various Data Science tasks that Spark Supports.

f) What do you understand by the term pair RDD?

g) With an example, explain the getNumPartitions() method.

h) What do you understand by the term checkpointing during Spark streaming?

i) List any four data types related to MLlib package.

j) Why is distinct() considered an expensive transformation?

PART - B

Answer any FOUR questions : **(6×4= 24)**

2. Differentiate between Transformations and Actions using a suitable example.
3. Compare and contrast the role of Accumulator and Broadcast variables in Spark.
4. Explain the benefits of running Spark on YARN.
5. Differentiate between Push-based and Pull-based receivers with an illustration.
6. Discuss the three methods that need to be implemented when implementing custom partitioners.
7. Write the steps that occur when running a Spark application on a cluster.

PART - C

Answer any THREE questions : **(10×3= 30)**

8. Highlight any 10 features of ApacheSpark.
9. Consider two dataframes df1 containing: Employee_ID, Employee Name, Department, Age and df2 containing: Employee_ID, Department, Designation, Location. Using python code demonstarte the working of inner join, outer join, left outer join and right outer join.
10. Discuss the four sections that the Spark UI provides for information and performance metrics as applications execute.
11. Describe the working of Apache spark with the help of an illustration.
12. Discuss the tips to be considered for improving the performance of MLlib in Spark applications.

****

19BDH302 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. THIRD SEMESTER DEGREE EXAMINATION FEBRUARY 2022

BIG DATA ANALYTICS

Artificial Intelligence and Deep Learning

Duration:3 Hrs **Max Marks:70**

PART - A

1. Answer any EIGHT of the following: **(2×8= 16)**

a) What is the fundamental concept followed in bidirectional search strategy?

b) What is the basic difference between Offline search and Online search algorithms?

c) Show from first principles that $P(a \mid b \wedge a)=1$.

d) What are the two quantities in the Bayesian learning approach?

e) Write the following sentences as Assertions in First Order Logic.
1. John is a King.
2. Richard is a person.
3. All Kings are persons.

f) What do you understand by the term proposition in propositional logic?

g) What is the fundamental logic of Markov chain Monte Carlo (MCMC) algorithms?

h) What is the purpose of memo functions in Explanation-Based Learning?

i) Distinguish between an Agent Fuction and an Agent Program.

j) What do you understand by the term Hidden variables?

PART - B

Answer any FOUR questions : **(6×4= 24)**

2. Summarize the PEAS discription for the following:
1. Part Picking Robot.
2. Refinery Controller.
3. Interactive English Tutor.

3. Discuss the rules of Universal Instantiation and Existential Instantiation.

4. Write the pseudocode for a decision-theoretic agent that selects rational actions.

5. Discuss any two applications of Reinforcement Learning.

6. Discuss the purpose of Hidden Markov Models? Construct a hidden Markov model to represent the process of predicting whether someone will be found to be walking, shopping or cleaning on a particular day depending upon on whether the day is rainy or sunny. Use suitable assumptions were required.

7. Differentiate between Active learning and Passive learning.

PART - C

Answer any THREE questions : **(10×3= 30)**

8. Write the psuedocode for Recursive best-first search.

9. Describe the fundamental concept followed in each of the following algorithms:
   1. Stochastic hill climbing
   2. First-Choice hill climbing
   3. Random restart hill climbing
   4. Local Beam search.

10. Describe in detail the four basic inference tasks that must be solved in temporal models.

11. Explain the working of Candidate Elimination algorithm with a suitable example.

12. Write the pseudocode of the algorithm for calculating minimax decisions. Explain its working with a suitable example.

****

19BDS304 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. THIRD SEMESTER DEGREE EXAMINATION FEBRUARY 2022

BIG DATA ANALYTICS

Advanced Machine Learning and NLP

Duration:3 Hrs **Max Marks:70**

PART - A

1. Answer any EIGHT of the following: **(2×8= 16)**

a) Write a python NLTK program to remove all numbers from a text.

b) Define probabilistic approach of parsing.

c) List out some applications of the question answering system.

d) List out the unsupervised learning algorithms.

e) Define Lemmatization.

f) Define statistical machine translation.

g) What is text classification used for?

h) Define Machine Learning.

i) Write a python NLTK program to get only nouns from a given sentence.

j) Define a shift-reduce parser.

PART - B

Answer any FOUR questions : **(6×4= 24)**

2. Compare information retrieval and boolean retrieval.
3. Explain chunking with an example program.
4. Explain Decision tree algorithm.
5. What is a web crawler and how does it work?
6. What are dictionaries?How do you create dictionaries in Natural Language Processing?
7. Write a program to perform dependency parsing using stanford parser.

PART - C

Answer any THREE questions : **(10×3= 30)**

8. Describe Machine learning based tagger.
9. Demonstrate dependency parsing.
10. Explain optical character recognition with an example.
11. Explain logistic regression with an illustration.
12. Explain the algorithm to perform question answering system.

****