

M.Sc DEGREE EXAMINATION, APRIL 2019
I Year II Semester
Data Science using Python

Time : 3 Hours

Max.marks :75

Section A ($10 \times 2 = 20$) Marks

Answer any **TEN** questions

1. What are the libraries present in scipy package?
2. Write the various operations performed on sets datatype.
3. What is a dataset?
4. How do impute missing data?
5. What are histograms?
6. Define hypothesis in data analysis.
7. What is SVD?
8. What are outliers in a data?
9. When is Gaussian distribution used?
10. Write the formula for recall in data accuracy.
11. How to find the skewness in a given data?
12. Define slicing term.

Section B ($5 \times 5 = 25$) Marks

Answer any **FIVE** questions

13. State the functionalities involved in machine learning concepts.
14. Discuss the interaction of data in NoSQL database.
15. Write about data patterns involved in scatterplots.
16. Discuss about Numpy package with example.
17. Discuss about the principal component analysis method?
18. Differentiate supervised and unsupervised learning methods.
19. What is the use of logistic regression? When it is applied?

Section C ($3 \times 10 = 30$) Marks

Answer any **THREE** questions

20. Discuss about iterator and dictionaries of python language in detail.
21. How to identify the redundant variables and missing values in a given data with example?
22. Discuss in detail about matplotlib lib function with proper examples.
23. Discuss about heuristic methods in clustering concept in detail?
24. Explain the K nearest Neighbour algorithm in fitting to a model?

M.Sc DEGREE EXAMINATION, APRIL 2019
I Year II Semester
Data Science using Python

Time : 3 Hours

Max.marks :75

Section A ($10 \times 2 = 20$) Marks

Answer any **TEN** questions

1. What are the libraries present in scipy package?
2. Write the various operations performed on sets datatype.
3. What is a dataset?
4. How do impute missing data?
5. What are histograms?
6. Define hypothesis in data analysis.
7. What is SVD?
8. What are outliers in a data?
9. When is Gaussian distribution used?
10. Write the formula for recall in data accuracy.
11. How to find the skewness in a given data?
12. Define slicing term.

Section B ($5 \times 5 = 25$) Marks

Answer any **FIVE** questions

13. State the functionalities involved in machine learning concepts.
14. Discuss the interaction of data in NoSQL database.
15. Write about data patterns involved in scatterplots.
16. Discuss about Numpy package with example.
17. Discuss about the principal component analysis method?
18. Differentiate supervised and unsupervised learning methods.
19. What is the use of logistic regression? When it is applied?

Section C ($3 \times 10 = 30$) Marks

Answer any **THREE** questions

20. Discuss about iterator and dictionaries of python language in detail.
21. How to identify the redundant variables and missing values in a given data with example?
22. Discuss in detail about matplotlib lib function with proper examples.
23. Discuss about heuristic methods in clustering concept in detail?
24. Explain the K nearest Neighbour algorithm in fitting to a model?