Tribhuvan University
Institute of Science and Technology
2078

✡

Master Level / I Year /First Semester/ Science
Data Science (MDS 501)
(Fundamentals of Data Science)

Full Marks: 45
Pass Marks: 22.5
Time: 2 hours

Candidates are required to give their answers in their own words as for as practicable.

Attempt All Questions

Group A

(5×3=15)

1. List and explain in short three main limitations of Data Science.
2. List three main differences between TDSP lifecycle and OSEMN framework.
3. List the differences between supervised and unsupervised machine learning methods including examples.
4. Define and briefly describe neural networks. List a few example use-cases where neural networks can be used.
5. What are decision trees? What does impurity of a node mean in context of decision tree?

Group B

(5×6=30)

6. Explain CRISP-DM in detail with a diagram and an example.

OR

What are the benefits of using OSEMN framework over CRISP-DM? In what scenarios will you use OSEMN instead of CRISP-DM?

7. Define with an example the K-means along with its limitations.
8. List major differences between linear and logistics regressions with examples.
9. Explain map-reduce programming paradigm. Include details on how it applies to Hadoop ecosystem.
10. Describe in detail how the quality of data can be assessed during data munging.

OR

List and describe how you would address various kinds of issues during data cleanup while doing data munging.

Tribhuvan University
Institute of Sciences and Technology
SCHOOL OF MATHEMATICAL SCIENCES
First Assessment 2078

Subject: **Fundamentals of Data Science** | *Full Marks*: **45**
Course No: **MDS 501** | *Pass Marks*: **18**
Level: **MDS First Year First Semester** | *Time*: **2 hrs**

Candidates are required to give their answer in their own words as far as practicable.

Attempt *ALL* Questions.

Group A **[5 × 3 = 15]**

1. Describe the applications and limitations of Data Science.
2. What is data science lifecycle? Briefly explain two major data science life cycles used by industries.
3. List and highlight the differences between structured, unstructured, and semi-structured data with examples of each.
4. Briefly explain the various methods used to handle missing values during data cleanup.
5. You want to identify global weather patterns that may have been affected by climate change. To do so, you want to use machine learning algorithms to find patterns that would otherwise be imperceptible to a human meteorologist. Discuss what machine learning method (supervised, unsupervised, reinforcement) would you use and why.

Group B **[5 × 6 = 30]**

6. Explain CRISP-DM, its advantages, and the steps involved in it.

OR

Explain TDSP lifecycle and the steps involved. Mention at least two main advantages of using TDSP lifecycle.

7. With an example, explain how you would determine the True Negative and False Negative data from researchdataset.
8. Explain with examples observation bias and funding bias in a research survey.
9. Describe, at a high-level, the major steps that need to be taken for data cleanup/munging.
10. Describe and explain with examples the various types of machine learning methods (Supervised, Unsupervised, and Reinforcement).

OR

Explain with examples and highlight the relationship between Artificial Intelligence and Machine Learning

***