

RAMAKRISHNA MISSION VIDYAMANDIRA

(Residential Autonomous College affiliated to University of Calcutta)

B.A./B.Sc. FIFTH SEMESTER EXAMINATION, MARCH 2022

THIRD YEAR [BATCH 2019-22]

COMPUTER SCIENCE (HONOURS)

PAPER : DSE1

Date : 03/03/2022

Time : 11 am – 1 pm

Full Marks : 50

Answer **any five** questions of the following:

[5×10]

1. a) Write down the difference between AI and ML? Briefly discuss about Machine Learning life cycle.
b) What is the difference between epoch and iteration? Briefly discuss about K-Nearest Neighbours (KNN) algorithm with a proper example. [(2+3)+(1+4)]
2. a) Briefly discuss about data pre-processing steps in ML.
b) What is the difference between Linear and Logistics Regression? What do you mean by Scaling, normalization in ML? [5+(3+2)]
3. a) Explain the Principle Component Analysis (PCA) algorithm with an example.
b) What is the need for feature engineering in ML? What is the limitation of cross validation? [5+(3+2)]
4. a) Define the terms: Hit Rate, False Positive Rate (FPR), Critical Success Index (CSI), Fowlkes-Mallows Index (FM). What is ensemble learning and why we use them in ML?
b) Briefly discuss the Decision Tree algorithm with an example and proper derivation. [(0.5+0.5+0.5+0.5+3)+5]
5. a) What are the different types of variables present in Data Science? Discuss the Ordinal encoding and One-Hot encoding technique using proper example.
b) Explain the K-Means Clustering algorithm with an example. [(2+1.5+1.5)+5]
6. a) Explain the different steps of Gradient descent along with proper derivation.
b) What is outlier? How will you deal with these outliers? [5+(2+3)]
7. a) Briefly discuss about Naïve Bayes Classifier algorithm with proper example. What is the application of Naïve Bayes Classifier?
b) Explain the needs of Heatmap, Confusion Matrix, Histogram, Whisker plot and Box plot for data visualization. [(4+1)+(1+1+1+1+1)]

————— × —————