

BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: MTECH (COGNIZANT)
BRANCH: CS

SEMESTER: I
SESSION : MO/2022

SUBJECT: CS537 SUPERVISED LEARNING

TIME: 3 Hours

FULL MARKS: 50

INSTRUCTIONS:

1. The question paper contains 5 questions each of 10 marks and total 50 marks.
 2. Attempt all questions.
 3. The missing data, if any, may be assumed suitably.
 4. Before attempting the question paper, be sure that you have got the correct question paper.
 5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.
-

- | | CO | BL | | | | | | | | | | |
|--|------|------|------|------|------|--------------|----|----|----|----|--|--|
| Q.1(a) What do you understand by a measure of central tendency and measure of dispersion? In case of a dataset containing many outlier which measure of central tendency shall not be a good choice and why? [5] | 1 | 2 | | | | | | | | | | |
| Q.1(b) Which of the following are examples of supervised learning. Give your reasons [5] | 1 | 3 | | | | | | | | | | |
| (a) Filtering spam mails from ham mails. | | | | | | | | | | | | |
| (b) Deciding the power of a person's spectacles. | | | | | | | | | | | | |
| (c) Detecting a speeding car at a toll gate. | | | | | | | | | | | | |
| (d) Putting plum cakes in a box, and pastries in a tray. | | | | | | | | | | | | |
| (e) Deciding whether a painting is by Van Gogh or Picasso. | | | | | | | | | | | | |
| Q.2(a) Why is accuracy often considered to be an insufficient metrics to judge the efficiency of a classification method. Suggest some other metrics which could be used instead. [5] | 2 | 2 | | | | | | | | | | |
| Q.2(b) Differentiate between a train, test and validation set. Why is it beneficial to shuffle a data set before splitting it into a train, test and validation set. Explain with an example. [5] | 2 | 3 | | | | | | | | | | |
| Q.3(a) Differentiate between an ordinal and a nominal attribute. Suppose we decide to encode three colours i.e. Red, Green and Blue using the codes 1,2 and 3. Do you think this is a good choice. If not why? [5] | 3 | 2 | | | | | | | | | | |
| Q.3(b) A classification algorithm is found to be overfitting a dataset. Do you think the algorithm has high bias or high variance. Why? [5] | 3 | 3 | | | | | | | | | | |
| Q.4(a) The winning times in the 100 mts race versus the year is given in the table below. Calculate the predicted winning time in 2024 using the technique of linear regression. [5] | 4 | 3 | | | | | | | | | | |
| <table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"><thead><tr><th>Year</th><th>2014</th><th>2016</th><th>2018</th><th>2022</th></tr></thead><tbody><tr><td>Time (secs.)</td><td>16</td><td>16</td><td>14</td><td>12</td></tr></tbody></table> | Year | 2014 | 2016 | 2018 | 2022 | Time (secs.) | 16 | 16 | 14 | 12 | | |
| Year | 2014 | 2016 | 2018 | 2022 | | | | | | | | |
| Time (secs.) | 16 | 16 | 14 | 12 | | | | | | | | |
| Q.4(b) Define the term entropy. Find the entropy of a dataset containing 10 individuals whose genders are {"Male", "Male", "Female", "Male", "Male", "Female", "Male", "Female", "Female", "Female"}. What does the result tell us? [5] | 4 | 2 | | | | | | | | | | |
| Q.5(a) Provide the statement of the Bayes theorem of probability w.r.t to classification. Why is the Laplacian correction often required while using the Naïve Bayes Classification? [5] | 5 | 2 | | | | | | | | | | |
| Q.5(b) What do you understand by an Ensemble Learning system. Explain with an example [5] | 5 | 2 | | | | | | | | | | |

:::17/03/2023 M:::