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

**CLASS: MTECH** **SEMESTER : II**
**BRANCH: AIML** **SESSION:SP/2023**

**SUBJECT: CS632 DATA ANALYTICS AND INTERPRETATION**

**TIME: 03 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 |
|---|---|---|---|
| Q.1(a) | If Z is a standard normal random variable, what is the probability that<br>(i) Z exceeds .62?<br>(ii) Z lies in the interval (-1.40, 1.40)?<br>(iii) \|Z\| exceeds 3.0?<br>(iv) \|Z\| is less than 2.0? | [5] | 1,1,1 |
| Q.1(b) | Using a t-Test to Confirm the Water Is Safe or not?<br>A city health department wishes to determine if the mean bacteria count per unit volume of water at a lake beach is within the safety level of 200. A researcher collected 10 water samples of unit volume and found the bacteria counts to be<br>**175, 190, 205, 193, 184, 207, 204, 193, 196, 180**<br>Do the data strongly indicate that there is no cause for concern?<br>Test with a α= .05.( From the t table, $t_{.05}$ with d.f. 9 is 1.833) | [5] | 1,3,1 |
| Q.2(a) | Water collected in a single bottle from a river is divided into eight specimens. Two specimens are randomly selected and sent to Lab A, two to Lab B, two to Lab C, and two to Lab D. The amount of heavy metals (ppm) is measured for each specimen. Calling each lab a treatment, suppose the data are<br>**Treatment — Observations**<br>**A: 5, 9**<br>**B: 8, 4**<br>**C: 4, 2**<br>**D: 7, 9**<br>(i) Obtain the arrays that show a decomposition for the observations.<br>(ii) Find the sum of squares for each array.<br>(iii) Determine the degrees of freedom for each sum of squares.<br>(iv) Summarize by an ANOVA table. | [5] | 2,4,2 |
| Q.2(b) | i) When is Ridge regression favourable over Lasso regression<br>ii) For Ridge Regression and LASSO, if the regularization parameter = 0, what does it mean?<br>iii) *What's the penalty term for the Ridge and Lasso regression?*<br>iv) What are the limitations of Lasso Regression?<br>v) For Ridge Regression and LASSO, if the regularization parameter is very high, what does it mean? | [5] | 2,2,2 |
| Q.3(a) | Explain how does an LDA model use for predictions? | [5] | 3,2,3 |
| Q.3(b) | Explain Why are SVMs often more accurate than logistic regression?<br>Give some situations where you will use an SVM over a CART Machine Learning algorithm and vice-versa. | [5] | 3,4,3 |

**PTO**

| | | | |
|---|---|---|---|
| Q.4(a) | How does a Random Forest calculate probability? How do I stop Overfitting in the Random Forest?<br>What is Boosting? Discuss with neat relevant examples. | [5] | 4,2,4 |
| Q.4(b) | Explain Deep learning and its architecture from an image-processing perspective. | [5] | 4,5,4 |
| Q.5(a) | Explain challenges in big data in Hierarchical and K-Means Clustering. | [5] | 5,3,5 |
| Q.5(b) | Explain the Q function and Q Learning Algorithm in Reinforcement Learning. | [5] | 5,6,5 |

**:::::19/07/2023:::::**