



Reg. No.

Manipal Institute of Technology

MANIPAL

(A constituent unit of MAHE, Manipal)

VII SEMESTER B. TECH. (COMPUTER SCIENCE AND ENGINEERING)

MAKEUP EXAMINATIONS, December 2019

SUBJECT: ELECTIVE V – MACHINE LEARNING WITH BIG DATA [CRA 4007]

REVISED CREDIT SYSTEM

(/12/2019)

Time: 3 Hours

Max. Marks: 50

Instructions to Candidates:

- Answer **ALL** questions & missing data may be suitably assumed.

- | | |
|--|----|
| 1.A. List and explain steps in machine learning process. | 5M |
| 1.B. Map the phases of CRISP-DM with machine learning steps. | 3M |
| 1.C. Explain with examples how missing values are handled in machine learning. | 2M |
| 2.A. With the help of examples explain different summary statistics. | 5M |
| 2.B. Explain data quality issues in machine learning. | 3M |
| 2.C. With the help of example explain recoding of feature. | 2M |
| 3.A. Explain how decision tree is computed in spark. | 5M |
| 3.B. Differentiate between classification and Regression. | 2M |
| 3.C. When overfitting and underfitting occurs in classification and what are their results? How to avoid overfitting? | 3M |
| 4.A. Why Validation Set is used? What are the several ways to create and use the validation set? Explain repeated holdout method. What are the problems with repeated holdout method and how this can be improved? | 5M |
| 4.B. What is the goal of linear regression? How the regression line can be determined? | 2M |
| 4.C. Explain cluster analysis. What are the different metrics used to measure the similarity between two samples? | 3M |
| 5.A. What are the different phases of classification model? | 2M |

- 5.B. What does the value of k indicate in k -means algorithm? Explain the different approaches for choosing the value of k ? 3M
- 5.C. What are the different types of errors in classification depending on the predicted and true labels? 5M
Explain the various metrics that can be used to evaluate the performance of classification model.