Reg. No. \_\_\_\_\_\_\_\_\_\_\_\_\_



**End Semester Examination – Nov / Dec – 2019**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Code :** | **18CS3031** | **Duration :** | **3hrs** |
| **Sub. Name :** | **DATA SCIENCE** | **Max. Marks :** | **100** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q. No.** | **Sub Div.** | **Questions** | **Course Outcome** | **Marks** |
| **ANSWER ANY FIVE QUESTIONS (5 x 16 = 80 Marks)** | | | | |
| 1. | a. | What is Data Science? Describe it with the Drew Convey’s Venndiagram of data science. | CO1 | 8 |
| b. | Discuss the various applications of data science. | CO1 | 8 |
|  |  |  |  |  |
| 2. | a. | Linear regression model is not a suitable filtering spam. Why? Discuss in detail. | CO2 | 8 |
| b. | Illustrate the Naïve Bayes classifier with an example of article classification. | CO2 | 8 |
|  |  |  |  |  |
| 3. | a. | Discuss in detail about the problems with nearest neighbhor’s approach. | CO3 | 8 |
| b. | Explain the various guidelines to build good production models. | CO6 | 8 |
|  |  |  |  |  |
| 4. | a. | Don’t prune the trees in random forests, because a great feature of random forests is that they can incorporate idiosyncratic noise. Illustrate the above point with an example. | CO3 | 12 |
| b. | What do you mean by over-fitting? Demonstrate how to handle this issue. | CO3 | 4 |
|  |  |  |  |  |
| 5. |  | What do you mean by dimentionality problem? Explain any two techniques which is used to handle this problem. | CO4 | 16 |
|  |  |  |  |  |
| 6. | a. | What are the two aspects to wrappers that you need to consider? Explain the different selection criterion in detail. | CO4 | 10 |
| b. | Give a detailed note on performance metrics that can be used to evaluate the performance of the various classifiers. | CO6 | 6 |
|  |  |  |  |  |
| 7. | a. | “Feature Generation or Feature Extraction”. Elucidate the different techniques that are available. | CO4 | 12 |
| b. | How do companies extract meaning from the data they have? Explain. | CO6 | 4 |
| **COMPULSORY QUESTION (1 x 20 = 20 Marks)** | | | | |
| 8. | a. | Describe various visualization techniques with fraud detection data. | CO5 | 10 |
| b. | Discuss in detail about the recent trends in various data collection and analysis techniques. | CO5 | 10 |