INSTITUTE OF AERONAUTICAL ENGINEERING
(AUTONOMOUS)

Code No: **BCS212**

MODEL QUESTION PAPER - II

M. Tech I Semester Regular Examinations

BIG DATA ANALYTICS

(Computer Science and Engineering)

Time: 3 hours Max. Marks: 70

Answer ONE Question from each Unit
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

UNIT I

1. (a) Define Big Data analytics and what made it powerful? What are the benefits of the Master Data Management (MDM) solution? [7M]
   (b) Suggest a solution when a salesman offers you a choice of three boxes, one containing a million dollars and two containing fifty dollars and tells you to pick one. He then shows you fifty dollars in one of the other two boxes and asks you if you want to change your choice to the remaining box that you have neither picked nor seen inside. [7M]

2. (a) Quote one of the biggest mistakes companies can make when starting an Master Data Management MDM project? [8M]
   (b) List out the steps that are to be taken reasonably in modeling when a model has to be built for a dataset of 200 patients with 4,000 variables including an indicator of whether or not they had developed cancer in the past year. [6M]

UNIT II

3. (a) Determine the advantages of Big Data analytics strategy which are often defined by the three V's -- volume, variety and velocity with examples. [8M]
   (b) Summarize the way using Big Data analytics when a experiment has to be run to see whether there is a difference in length of visit between customers viewing your old website and your new website [6M]

4. (a) Describe the method by which companies analyze customer data or other types of data in an effort to identify patterns and discover relationships between different data elements. [7M]
   (b) State the reason for the problem and procedure to overcome , when an attempt is done to model the price of cars at auction and you find that your model has trained well, but subsequently does poorly on new data. [7M]

UNIT III

5. (a) What are the stages in Map Reduce? Explain pair design patter on a co-occurrence example. Include advantages and disadvantages against stripes approach, possible optimizations and their efficacy. [6M]
   (b) Explain the installation process of HDFS and explain one of the best ways to assess a model usefulness which is built when marketing department of a company is looking for a way to call customers who are likely to churn and persuade them to stop. [8M]

6. (a) Describe the reasons for recommending Hadoop as a replacement of enterprise data warehouse. [7M]

(b) How do you decide which parts of the process are amenable to parallelism, when you were asked to construct an ensemble model but are not sure about the best way to go about it. [7M]

UNIT IV

7. (a) List out the writable data types used by Hadoop API that have the same features as default java classes. [6M]

(b) Explain about the Hadoop input and output with a example and write a note on data integrity? [8M]

8. (a) Discuss Hadoop streaming with RUBY and PYTHON programming language with at least one example. [7M]

(b) Can Hadoop be used to create distributed clusters, based on commodity servers, which provide low-cost processing and storage for unstructured data, log files and other forms of Big Data explain with suitable example. Express. [7M]

UNIT V

9. (a) Describe the future of sentiment analysis .List out the techniques used in sentiment analysis. What are the measuring behaviors in mobile analytics? [7M]

(b) Which of the tasks do you prioritize for a model which is first built to run on a small data but later has to be used to run for millions of customer data. [7M]

10. (a) Explain sentiment classification using supervised learning .Your CEO is worried that your company is not very responsive to social media and wants a data science-driven way of fixing this. Which options do you suggest? [7M]

(b) What are the applications for mobile analytics? List the different ways or strategies to measure the success of mobile application. [7M]