****

Register Number:

DATE: 21-04-2017

**ST. JOSEPH’S COLLEGE (AUTONOMOUS), BENGALURU-27**

**M.Sc. BIG DATA ANALYTICS – II SEMESTER**

**SEMESTER EXAMINATION: APRIL 2017**

**BDA -DE-2516: Multivariate Statistics**

**Time 2.5 Hours Maximum Marks 70**

**This Question Paper Contains TWO Printed Paper And ONE Part**

**Answer AS MANY QUESTIONS AS POSSIBLE 7 x10 = 70**

1a. State any three axioms of probability [for example, P(sure event) = 1]. |6|

1b. If P(A) = 1/3 and P (B) = 2/3, then what is P (AUB) if:

* A and B are independent |3|
* A and B are disjoint |3|

2. Let X be your expected mark in this exam. Let Y be the number of hours that you studied for this exam. Create a dummy X-Y data set for 10 students and then:

* + Compute the correlation coefficient between X and Y |5|
  + Draw the scatter plot of X vs Y |2|
  + Write down the regression equation of Y (dependent variable) on X |5|

BDA-DE-2516-A-17

3. Discuss how you can convert the bivariate problem of Question 2 into a multivariate problem. Specifically highlight the following points (don’t write more than one page in all)

* New independent variables you might add |4|
* Probable presence of collinearity |4|
* Using R squared, or adjusted R squared? Which one? Why? |4|

4a. Describe (in no more than 5 sentences) the benefits of principal component analysis |5|

4b. Sketch (as a flow chart) the different steps involved in PCA |5|

4c. Mention one application where PCA can make a big difference |2|

5a. What is the underlying principle of clustering? |4|

5b. Give two real-life examples (from sport of business) where cluster analysis helps |4|

5c. Sketch (as a flow chart) the rationale of k-means clustering |6|

6. A bank has a tricky decision to make. Should it offer a credit card to a customer with a seemingly modest income?

* What sort of multivariate analysis would you recommend? |2|
* List out 10-12 possibly predictive variables? |6|
* Sketch (as a flow chart) your options of stepwise regression |6|

7. Write short notes on any two of the following: |7+7|

* + Solving classification problem by clustering
  + Multivariate techniques in social media
  + The promise of artificial intelligence