SRN | | | | | | | | | | | | |

# PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

**UE16CS521**

## JAN 2017: END SEMESTER ASSESSMENT (ESA) M.TECH. I SEMESTER

## UE16CS521- DATA ANALYTICS

Time: 3 Hrs | Answer All Questions | Max Marks: 100

| | | | |
|---|---|---|---|
| 1. | a) | Suppose you are given the names of 600,000 villages in India. You are given just the names and the state where they are located. No other data about the villages is available. Can you do data analytics with this data set? What types of analytics can you do? Propose at least 2 examples of interesting analytics that you can perform with this data set. | 2+4 +4 |
| | b) | Consider the statement: "Data analytics is nothing but statistics." Do you agree with this? Justify your answer. | 5 |
| | c) | Consider the statement: "Data analytics is nothing but machine learning." Do you agree with this? Justify your answer. | 5 |
| | | | |
| 2. | a) | State Chebyshev's Theorem. | 4 |
| | b) | How is it used in practice? | 4 |
| | c) | What is Chi-Square test? How is it used? | 4 |
| | d) | What is T-Test? How is it used? When do we use which test? | 8 |
| | | | |
| 3. | a) | In Time-Series Analysis, what are the differences between Successive Differences, Moving Average and Auto-correlation? Illustrate each with an example to highlight the differences and the uses of each of the three techniques. | 6+6 |
| | b) | Suppose a time-series data has a linear trend with a positive slope *m*. How will you de-trend it? | 4 |
| | c) | Suppose a time-series data is said to have a seasonality with a period of 6 months. How do you verify that there is indeed a seasonality with that periodicity? | 4 |
| | | | |
| 4. | a) | In a single-server queue, what is the expected length of the queue if (i) arrival rate is half the exit rate? (ii) arrival and exit rates are the same? (iii) exit rate is zero? (iv) arrival rate is zero? | 4 |
| | b) | If we have two servers instead of one, with the same arrival rate, what happens to the length of a single queue serviced by both servers? | 4 |
| | c) | Give an example of a Partitional Clustering Algorithm. Write its pseudocode. Illustrate how it works. | 2+4 +4 |
| | d) | Write the formula for Minkowski Distance. | 2 |
| | | | |
| 5. | a) | What is the difference between reliability and validity of a classification? | 4 |
| | b) | What are Risk Ratio and Odds Ratio? What is the difference between them? | 6 |
| | c) | What is logistic regression? What is its formula? When do you use it? | 6 |
| | d) | Is this true or false? Why? "If y is related to x exponentially, it is still possible to use linear regression to predict the value of y for a given value of x." | 4 |