

**Time : 2 Hours****Total marks : 60**

- N.B. (1) All questions are compulsory.  
 (2) Figures to the right indicate full marks.  
 (3) Assume additional data if necessary but state the same clearly.  
 (4) Symbols have their usual meanings and tables have their usual standard design unless stated otherwise.  
 (5) Use of calculators and statistical tables are allowed.

<b>Q1</b>	<b>Attempt Any two of the following.</b>	<b>12</b>
a	Explain the Partitioning Algorithm for a large data set, with example?	6
b	Explain the Hierarchical Algorithms: Agglomerative (AGNES), Divisive (DIANA).	6
c	Discuss KNN in detail.	6
d	What is big data? Discuss challenges of big data.	6
<b>Q2</b>	<b>Attempt Any two of the following.</b>	<b>12</b>
a	Explain Confusion matrix in detail..	6
b	Write a note on a) Precision b) Recall	6
c	Explain in detail CHAID.	6
d	Give the difference between Regression and Classification.	6
<b>Q3</b>	<b>Attempt Any two of the following.</b>	<b>12</b>
a	Explain the concept of PageRank	6
b	Describe Spam Mass.	6
c	Explain the PageRank iteration using MapReduce.	6
d	Write a note on Singular value decomposition.	6

**Q4      Attempt Any two of the following.      12**

- a      Describe Clustering users and items      6
- b      Describe the duality of similarity.      6
- c      Explain Collaborative Filtering.      6
- d      Illustrate the applications of recommendation system.      6

**Q5      Attempt Any two of the following.      12**

- a      Describe Clustering users and items.      6
- b      Explain the interpretation of SVD.      6
- c      Discuss how you will use eigen vectors for dimensionality reduction.      6
- d      What do you mean by Item profiles?      6

\*\*\*\*\*