

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- March-April 2017

M.Tech IV Semester

COURSE CODE: 11M1WCI431

MAX. MARKS: 25

COURSE NAME: Advanced Web Mining

COURSE CREDITS: 03

MAX. TIME: 1.5 Hrs

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

- Doc 1 new home sales top forecasts
- Doc 2 home sales rise in july
- Doc 3 increase in home sales in july
- Doc 4 july new home sales rise
- Doc 5 breakthrough drug for schizophrenia
- Doc 6 new schizophrenia drug
- Doc 7 new approach for treatment of schizophrenia
- Doc 8 new hopes for schizophrenia patients
1. Using cosine similarity and k-means algorithm cluster the documents into 2 groups. [8]
 2. Consider the documents {1, 2, 3, 5, 6,7} as training documents classify {4, 8} using naïve Bayes algorithm. [10]
 3. What is boolean indexing and inverted indexing? Explain map-reduce algorithm for constructing a large scale inverted index. [7]