Subject:	**Data mining and Business Intelligence (**ITC602**)**		
Class:	**TE IT Semester – VI (CBCGS) Academic year: 2018-19 (FH 2019)**		
Name of Student:			
Roll No:		Date of performance (DOP) :	
Assignment No:	**Theory Assignment 02**	Date of checking (DOC) :	
Marks:		Teacher's Signature:	

Marks shall be evaluated as out of 20 (15 marks for answers, 03 marks for presentation and 02 marks for punctuality)

Batch: B1

Q.1		What is Jaccard co-efficient? How is it calculated? What is its application?	(03)
Q.2		What is the difference between agglomerative and divisive clustering? Give example. Why it is called as hierarchical clustering?	(03)
Q.3		What is over fitting and tree pruning? Explain with example	(04)
Q.4		Describe the decision tree algorithm.	(05)

Batch: B2

Q.1		Why clustering is called as unsupervised learning? Explain with example	(03)
Q.2		How the quality of good clustering can be described? Explain with example	(03)
Q.3		What is an ensemble method? Give example of ensemble method	(04)
Q.4		Describe the attribute selection measures used in Decision Tree algorithm	(05)

Batch: B3

Q.1		What is data matrix and dissimilarity matrix? Give one example.	(03)
Q.2		What is k-fold cross validation?	(03)
Q.3		What are the features of Bayesian classifier?	(04)
Q.4		Enlist different clustering approach.	(05)

Batch: B4

Q.1		Compare between bagging and boosting	(03)
Q.2		What is ROC curve? How it can be used to estimate the performance?	(03)
Q.3		Explain Bayes' Theorem. What is priori and posteriori probability?	(04)
Q.4		What is single link and complete link in clustering? Explain with example.	(05)