Digital Assessment

Subject: Web Mining (CSE 3024) **Slot : G1 + TG1**

Online Submission Deadline: 17th June 2020

Implement the web content mining techniques assigned against your registration name.

Note: Combine all the implemented python files into a single .ZIP file and upload it in **Schoology and upload the contents as well as output of all the files as a single PDF file in VTOP**. The uploaded material must contain the implemented code and snapshot of the output of each program **(both classification & clustering)**.

UPLOADING FILES IN BOTH SCHOOLOGY AND VTOP IS MANDATORY.

Don't forget to mention your group number in the uploaded submission file. The student needs to carry out the task **individually or in a group of 2 members only**. If the implementation has been carried out in a group, then in the report both the students name must be mentioned.

NO request of considering the DA for award of marks will be entertained once the deadline is over. If any student misses the deadline to upload the assignment on or before the due date, then a Zero mark will be awarded.

Please click the following link to download the dataset:

https://drive.google.com/open?id=1R1v1FzPL_hbIpSHOW5qMLh7DOuhvh-25

<table>
<tr><th>Regd. No.</th><th>Name</th><th>Dataset</th><th>Techniques to be Implemented</th></tr>
<tr><td colspan="4">Group 1</td></tr>
<tr><td>17BCE0118</td><td>GEORGE JOSEPH CHAKOLA</td><td colspan="2">Clustering</td></tr>
<tr><td>17BCE0127</td><td>NAMAN GUPTA</td><td rowspan="2">User Identify-Walking Activity.zip</td><td rowspan="2">1. K-Means Clustering
2. Probability-based Clustering</td></tr>
<tr><td>17BCE0332</td><td>S MAHISHA</td></tr>
<tr><td>17BCE0361</td><td>SUYASH SRIVASTAVA</td><td colspan="2">Classification</td></tr>
<tr><td>17BCE0367</td><td>D SRIVATHSSA</td><td rowspan="3">breast-cancer.zip</td><td rowspan="3">1. Decision Tree (C 4.5)
2. Random Forest
Compare the performance of classifiers using the following metrics:
a) F - Score b) ROC & AUC Curve
c) R^2 d) Error Rate e) Accuracy</td></tr>
<tr><td>17BCE0453</td><td>VISHAL LADHAR</td></tr>
<tr><td colspan="2"></td></tr>
</table>

Group 2			
17BCE0481	RONAK JAIN	**Clustering**	
17BCE0529	PONNADA SAI GOWTHAM BHARADWAJ	**Health-News-Tweets.zip**	**1. K-Means Clutering** **2. BIRCH Clustering**
17BCE0564	SHASHANK NISHAD		
17BCE0582	KEDARNATH K CHIMMAD	**Classification**	
17BCE0619	MAHESH SHARMA	**DelhiWeather.zip**	**1. Naïve Bayes** **2. Gradient Boosting** Compare the performance of classifiers using the following metrics: **a)** F - Score **b)** ROC & AUC Curve **c)** R^2 **d)** Error Rate **e)** Accuracy
17BCE0686	RAJAT R PANDURANGI		
Group 3			
17BCE0756	PRIYANSH AGRAWAL	**Clustering**	
17BCE0811	SHARAN KUMAR B CHINIWAR	**User Identify-Walking Activity.zip**	**1. Hierarchical Clustering (Use Manhattan Distance Function)** **2. Probability-based Clustering**
17BCE0870	YETURU JAGADEESH REDDY		
17BCE0954	ACHYUT TRIPATHI	**Classification**	
17BCE2067	SHOBHIT BISHNOI	**car-eval.zip**	**1. Decision Tree (C4.5)** **2. SVM** Compare the performance of classifiers using the following metrics: **a)** F - Score **b)** ROC & AUC Curve **c)** R^2 **d)** Error Rate **e)** Accuracy
17BCE2142	L DHINESH KUMAR		
Group 4			
17BCE2149	ANUJ JHUNJHUNWALA	**Clustering**	
17BCE2193	SUBHADITYA MUKHERJEE	**Health-News-Tweets.zip**	**1. Hierarchical Clustering (Use Euclidian Distance Function)** **2. Expectation Maximization**
17BCE2270	PRIYANK DEV VARSHNEY		
18BCE0265	ISHAAN OHRI	**Classification**	
18BCE0283	ANMOL PANT	**diabetes.zip**	**1. Decision Tree (C4.5)** **2. Adaboost** Compare the performance of classifiers using the following metrics: **a)** F - Score **b)** ROC & AUC Curve **c)** R^2 **d)** Error Rate **e)** Accuracy
18BCE0633	RAUNAQ NIJHAWAN		
Group 5			
18BCE0667	AISHA SARTAJ	**Clustering**	
18BCE0688	AKSHAT KASHYAP	**User Identify-Walking Activity.zip**	**1. Hierarchical Clustering (Use Manhattan Distance Function)** **2. Expectation Maximization**
18BCE0706	ANJALI ROY		

18BCE0720	ANANYA PATANKAR	**Classification**	
18BCE0724	ARCHIT AGGARWAL	**iris.zip**	**1. Naïve Bayes** **2. Adaboost** Compare the performance of classifiers using the following metrics: **a)** F - Score **b)** ROC & AUC Curve **c)** R^2 **d)** Error Rate **e)** Accuracy
18BCE2062	HARSHIL SHAH		
Group 6			
18BCE2117	SHLOK KOOLWAL	**Clustering**	
18BCE2180	SOHAM DEBROY	**Health-News-Tweets.zip**	**1. BIRCH Clustering** **2. Probability-based Clustering**
18BCE2395	ASHISH THAPA		
18BCE2432	EPHREM BEKELE DUBISSO		
18BCI0182	VIVEK MEHTA	**Classification**	
		heartdisease.zip	**1. Naïve Bayes** **2. KNN** Compare the performance of classifiers using the following metrics: **a)** F - Score **b)** ROC & AUC Curve **c)** R^2 **d)** Error Rate **e)** Accuracy