

Digital Assessment

Subject: Web Mining (CSE 3024)

Slot : C1 + TC1

Online Submission Deadline: 20th May 2021

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 **MS Team Assignment** 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 VTOP AND MS TEAM ASSIGNMENT SECTION 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 (from the same group)**. 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/file/d/1-kNS0Hnxm2iYmVUJDBwpi223P2RrH1Hj/view?usp=sharing>

Regd. No.	Name	Dataset	Techniques to be Implemented
<u>Group 1</u>			
18BCE0358	AADITYA PRAKASH SINGH	sales transc.zip	Clustering 1. K-Means Clustering 2. Probability-based Clustering
18BCE0391	PATEL DHRUV		
18BCE0409	NITISH SHARMA		
18BCE2121	JESWIN JACOB J	PD_Speech.rar	Classification 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) Accuracy
19BCE0300	MD HISHAAM AKHTAR		
19BCE0406	SOUMYA RANJAN TRIPATHY		
19BCE0498	AYUSH KHARE		
19BCE0504	ANOUSHKA DALAL		

Group 2			
19BCE0521	AMAN ANAND	Clustering	
19BCE0625	MUKUNTH B S	plants.zip	1. Hierarchical Clustering (Use Euclidian Distance Function) 2. BIRCH Clustering
19BCE0664	VIBIN BHARAT V N		
19BCE0674	DEVESH D R	Classification	
19BCE0682	K SUJIT KUMAR	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) Accuracy
19BCE0685	PRAVEEN RAJ M		
19BCE0744	AGNIVA BASAK		
19BCE0778	S NAYAN		

Group 3			
19BCE0837	YASH MEHTA	Clustering	
19BCE2036	SHIVAM M MEHTA	sales transc.zip	1. BIRCH Clustering 2. Probability-based Clustering
19BCE2037	JHANAK GUPTA		
19BCE2165	ARYAK ROY	Classification	
19BCE2305	AKSHITA KHALORIA	heartdisease.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) Accuracy
19BCE2311	GAURAV SINGH		
19BCE2393	MEHROTRA AYUSH VIPUL		
19BCE2451	PRAKHAR VYAS		

Group 4			
19BCE2524	ARYAN TODI	Clustering	
19BCE2525	ABINASH KUSHWAHA	plants.zip	1. Hierarchical Clustering (Use Euclidian Distance Function) 2. Expectation Maximization
19BCE2535	ISAN AGRAWAL		
19BCE2536	AYUSH RUNGTA	Classification	
19BCE2671	PRAGATI NATWARLAL BHATTAD	DryBeanDataset.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) Accuracy
19BCE2672	LOKESH MISHRA		
19BCE2678	VISHAKHA KUMARESAN		
19BCI0224	NEELAY MITRA		
19BEC0863	AYUSHI MISHRA		