***WORKSHOP ON WEB CRAWLING AND SCRAPING***

***Student Trainer: Ganish Deepak***

1. ***10AM - 11.15AM (\*MODULARIZED WEB SCRAPING) IS TOUGHT WITH PARALLEL IMPLEMENTATION IN COLAB / CLOUD MACHINE***
2. ***11.15AM – 11.25AM EXPLANATION A DEPLOYED AND FULLY FUNCTIONAL WEB SCARPING SOFTWARE WITH FLASK UI.***
3. ***11.25AM - 11.50AM STUDENTS WILL BE ASKED TO RECREATE THE EXPLAINED SOFTWARE WITH MINIMAL CONSOLE UI***
4. ***11.50AM – 12.00PM WEB SCRAPING ETHICS AND LEGAL CONSIDERATIONS + Q&A SESSION***

***(\*MODULERIZED WEB SCRAPING) - SYLLABUS***

1. ***INTRODUCTION TO WEB CRAWLING AND WEB SCRAPING***
2. ***SETTING UP PYTHON ENVIRONMENT WITH MODULES REQUIRED***
3. ***REQUEST LIBRARY: SENDING HTTP REQUEST TO RETRIVE HTML PAGES***
4. ***BEAUTIFULL SOUP LIBRARY: PARSING RETRIVED PAGES***
5. ***NAVIGATING AND SEARCHING HTML TREE STRUCTURE***
6. ***HANDLING PIGNATION AND FOLLOWING LINKS***
7. ***STORING SCRAPPED DATA IN A CSV OR JSON FILE***