WQD7009: Course Assignment (40%)

Deadline for Part 1 and Part 2 is on the 8th January, 2021 (0000hr)

Part 1: (30%)

a) Market Basket Analysis (15%)

Method of Evaluation: From Report

Perform Market Basket Analysis on the provided dataset. To provide a report containing the following sections:

- Introduction to Frequent Itemsets & Association Rules Mining.
  To include explanations on the followings:
  - Support
  - Confidence
  - Lift
  - Conviction
- Introduction to the dataset
  - Perform data Analysis & Data Exploration
- To extract the Association Rules from the dataset using A Priori algorithm using Python
  - Hint: Usage of mlxtend package available in Python

b) Using Singular Value Decomposition (SVD) in a Recommender System (15%)

Method of Evaluation: From Report

Study the tutorial provided in the following link:

https://analyticsindiamag.com/singular-value-decomposition-svd-application-recommender-system/#:~:text=In%20the%20context%20of%20the,given%20to%20items%20by%20users.

Based on the tutorial, write a report for Section (B) of Part (1) that contains the followings:

- Introduction to the different techniques involved
  - SVD
  - Recommender System
  - Collaborative Filtering
- Elaboration on the theoretical concept with specific examples taken from the tutorial with specific examples presenting the concepts behind
  - SVD

    - Recommender System
    - Collaborative Filtering
- Working demo with GUI using either
    - https://www.streamlit.io/ OR
    - https://blog.jupyter.org/and-voil%C3%A0-f6a2c08a4a93

Part 2:

Method of Evaluation: Presentation

Each group is required to select one of the papers below

- Random Forests for Big Data
  https://www.scopus.com/record/display.uri?eid=2-s2.0-85028618822&origin=inward&txGid=bcccf5b2ad30b347275b36565f669743
- A Novel Clustering Method Using Enhanced Grey Wolf Optimizer and MapReduce
  https://www.scopus.com/record/display.uri?eid=2-s2.0-85047493065&origin=inward&txGid=08fa63e9bcdcb41f272b61c8eea93227
- Frequent Itemsets Mining for Big Data: A Comparative Analysis
  https://www.scopus.com/record/display.uri?eid=2-s2.0-85028451362&origin=inward&txGid=8c6896c3ad07f555cfc6c97084233c71
- Train Delay Prediction Systems: A Big Data Analytics Perspective
  https://www.scopus.com/record/display.uri?eid=2-s2.0-85020215215&origin=inward&txGid=984dae39a178f31e3a42e3104aab8636

Based on the paper selected, each group needs to prepare a presentation slides containing the followings:

- Introduction
- Problem Statement
- Objectives
- Methodology
- Results
- Conclusion

** The evaluation will be solely based on the presentation and based upon the followings

- Clarity of the presentation and the paper presented
- Understanding of the paper (evaluated from Q&A)
- Creativity in presenting the paper