Data Science Exercise

Q1 – Collaborative Filtering for Social Follower Data

Given a SQL database with a schema of (row_id, user_id, following_user_id), write a short python code to retrieve the data from the SQL database, highlighting assumptions or code abstractions (e.g. login, password, db server address) and use any library of your choice to do collaborative filtering to recommend new users to follow.

Q2 – Flask web-service wrapper

Write some python code that makes use of Flask or another framework to expose your recommendation engine through a web service to external clients.

Q3 – User vectors

In content-based recommendation, the aim is to recommend other users based on vectors that describe the characteristics of users. (a) Describe what data would be useful to include in those vectors to give the best recommendations (b) Give some examples of distance metrics that can be used to compare these user vectors.