Capstone Project 2:

Academic Paper Recommender:

The aim of this project is to create a recommender system to help researchers and scientists find related articles to an specific paper. There are thousands of papers published every day and it is not feasible for a researcher to browse all these papers in order to find a related paper to a desired subject. This recommender system helps scientists with their search and saves them a lot of time looking into article resources.

The data for this project is collected using API for PubMed, which is a repository for biomedical data. The data is extracted in raw XML in a full text format and information such as paper's ID, title, authors last name, year, journal, abstract, tags, and citations is collected from each paper.

Different methods and information is used to come up with a recommender for a paper. The first approach is text similarity-based recommender. In this method, title and abstract of each paper is used and a TF-IDF based similarity is calculated in order to recommend the n number of related papers. The other method is based on the tags used for each paper. Papers which are sharing n number of tags with the reference paper would be recommended. The other similarity can be based on the citations, such that finding papers which have similar citations to our reference paper.

The deliverables for this project would be a Python notebook and a report describing the problem, the approach and the results.