Capstone project 2: Project Proposal

1. Title: Application Recommendation System for Biodegradable Polymer

2. Problem:

Developing new applications of a polymer is one of the most important and challenging steps to sell a polymer as a product. It requires a wide range of knowledge about polymer itself (physical property, formability, etc.) and products that might already exist or might not yet. Scientists would use research papers, databases, and patents as a reference. However, this is a quite painstaking process to search similar polymers' features and the applications, and combine all information to think about new applications of their polymer.

Here, I would like to provide an application recommendation system that gives some hints about potential applications for a given polymer by a user. I chose a biodegradable polymer as a polymer kind for this project. Biodegradable polymers have been researched for decades. There are mainly two aspects to get attention; to reduce the effects of plastics on the environment, and as a bioabsorbable polymer, which are degraded and absorbed in our body. Both fields are growing with raising environmental awareness and advancing in medical technology.

3. Client:

The first clients will be chemical companies that research and sell biodegradable polymers as raw material and that buy the polymers and form them for their products. Their purposes could be to accelerate the step to extract potential applications of biodegradable polymers.

4. Dataset:

Patents will be used to make the system. A patent includes information about what polymer was used, how to prepare it, for what it could be used, who invented it, and so on. The material and application information would provide useful information for this project.

The patent data will be acquired from the database of the World Intellectual Property Organization (WIPO) with a search word "IC:(C08L 101/16)". This data contains Application ID, Application Date, Country, Title, Abstract, IPC Codes, Applicants, Inventors. etc. C08B 101/16 is the IPC code, International Patent Classification Code, for biodegradable polymers. IPC has a nested structure, and each code represents a material, application, or technology to manufacture it.

5. Approach:

The first step will be to clean and transform the data. The necessary columns and the rows will be extracted. Also, this step includes identifying missing data, duplicate data, and outliers.

The second step will be to implement a network analysis by IPC codes. Generally, a patent has several IPC codes. A patent will be one kind of node, and an IPC code will be another kind of node. A patent-node will connect with some code-nodes (the number of connections depends on the number of IPC codes the patent has). Because the system

will have a lot of patents, the connections become a network form; a patent shares some code-nodes with other patents. This network will represent the relationships between patents and IPC codes.

On the third step, a system extracting IPC codes close to a given patent (or IPC codes) will be created. This step includes making a system to calculate a similarity between patents by their abstract texts. When this system is given a patent (or IPC codes) by a user, this system will extract the nearest patents to the given one using the network. Then, the nearest patents will be ordered by the similarity of their abstracts. Finally, the IPC codes of some most similar patents will be provided to the user as the potential applications.

6. Deliverables:

a. Codes (Jupyter Notebook)
b. Report
c. Presentation