LAWRENCE TECHNOLOGICAL UNIVERSITY
Department of Math and Computer Science

INT6303
Intro to Social Media Data Analytics
M. Al Hamando, Ph.D.

Guidelines for Project Proposal and Implementation

You are required to develop and submit a project involving the design and implementation of a social media data analysis-based application. Your project proposal should include the following:

Brief Overview:
- What is the general domain?
- Why is it important?

Problem Statement & Description:
- What is the general and specific problem that you are addressing?
- What are the high-level functionalities expected to be provided in this project?

Solution (initial ideas only):
- Based on the proposed solution, try to define initial specifications for your project.
- What is the scope of the problem/solution (what is included/excluded)?

Data:
- Source for data, size, scope, …

Proposed Project details and Timeline
- Steps you are planning to follow in order implement the project?
- Individual or team project? Project only or research paper combined?
- Using own laptop, a server, or the cloud?

Other:
- Any other specific issues to be addressed

The problem:

There is an increasing interest in the higher education administrators in the image of their institutions. This image is very important to the success of these institutions.

The solution:

The amount of data that is present in Social Networks provides an opportunity for the administrators to gain insights from this data that is being generated by students, alumni, families, communities, as well as prospective students. With the increasing volume of data that is being generated and with the advancements in data mining and machine learning fields, it is possible to capture these sentiments and classify them based on their attitudes as well as potentially making predictions about future decisions in terms of enrollment and retention. The application could also be able to detect communities of similar individuals and experiences.

The problem:

There has been a great interest in politics to measure the approval rating of politicians and institutions, e.g., Congress, Government, Supreme Court, and Law-enforcement agencies. Traditionally, some statistical methods have been used to measure these ratings, such as polls and surveys.

The solution:

The amount of data that is present in Social Networks provides an opportunity for the politicians, the media, as well as for the public to gain insights from this data that is being generated by individuals, organizations, and by communities. With the increasing volume of data that is being generated and with the advancements in data mining and machine learning fields, it is possible to capture these sentiments and classify them based on their attitudes as well as potentially making predictions about future decisions in terms of future elections or success in passing new legislations. The application could also be able to detect communities of similar individuals and expectations.

The problem:

While network structures are good representations of the relationships between nodes, they do not represent the communities that exist within the larger network. The purpose is to bring out more meaningful information about the networks. Moreover, finding overlapping community is a challenging problem.

The solution:

You need to explore methods to extract communities from a given network. In many networks, vertices may belong to more than one group, and such groups form overlapping communities. Some examples are social networks, where an individual usually belongs to different circles at the same time, such as work colleagues, family, sport associations, etc. The application can be used to identify target communities in the context of business requirements that focused communities that belong to specific categories such as Java Programming, Car Sales, Travel, Restaurant Business, and Football.

The problem:

The advent of Internet and Web 2.0 have resulted in the abundance of behavioral data about digital life. Digital life becomes a major part of daily life as people increasingly spend their time online interacting with each other in virtual places like Facebook, Twitter, Instagram, among others. In fact, various offline behavior previously unnoticed or unrecorded has now become more prominent as online technologies broadcast and record these behaviors. However, at the aggregate level, online behavior is often not linked to offline location based behavior.

One such case is cyberbullying. It is unclear whether cyberbullying is an aggressive behavior that will also exhibit offline in the same physical location.

The solution:

The project is a proposal to analyze social media-based cyberbullying activities across the U.S. as an example to understand the connection between placed-based social behavior and cyberspace-based behavior. This project can focus on one or more of the following types of aggressive cyberbullying behavior:

- traditional workplace bullying
- schools bullying
- adolescent bullying

Deliverables

- Proposal (Tuesday 6/20)
    - As described above
- Final application (Tuesday 7/18)
    - Functional Requirements
    - Code (as part of the documentation of the project) and data
    - Application (soft copy)
    - Documentation (minimum of 4 pages)
    - Presentation & Demo (in class)

Rubric and Evaluation Criteria for the project

Content (80%):

Complexity of the problem	5 pts
Definition and description of the problem	5 pts
Analysis of the problem	5 pts
Research and investigation	5 pts
Application design and architecture	10 pts
Completeness of solution (satisfy all requirements)	10 pts
Technical implementation	40 pts
Utilization of Data mining techniques	
Visualization of analyses and results	
Executes properly	

Documentation (10%):

Proper writing format	2 pts
Clarity of writing, structure, flow	3 pts
Good documentation of the approach/solution	5 pts

Presentation (10%):

Proper, flowing and coherent presentation	2 pts
Clarity of oral communications	3 pts
Quality of presentation materials (code, demo)	5 pts