CS 412 — Natural Language Processing
Assignment #3

Due Date: Sunday, Nov 22th 2020 by 11:59pm.

Assignments are to be done individually. No late assignments will be accepted. Submit your assignment by the deadline on the day that it is due. **Submissions that do not comply with the specifications given in this document will not be marked and a zero grade will be assigned.**
You are to submit the Complete Python notebook on LMS.

Basic Data Scrapping and Sentiment Analysis

1 Introduction

Your goal in this assignment is to implement and practice some basic understanding of Data Scrapping and Sentiment Analysis. Data scraping, also known as web scraping, is the process of importing information from a website into a spreadsheet or local file saved on your computer. It's one of the most efficient ways to get data from the web, and in some cases to channel that data to another website. Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. ... Sentiment analysis helps data analysts within large enterprises gauge public opinion, conduct nuanced market research, monitor brand and product reputation, and understand customer experiences.

2 Assignment Tasks

The first task that you should perform is to modify name of the notebook with your id and the first cell in the notebook to print your name without any vowels except the uppercase first letter of your first and last name (no credit will be given if this task is not completed correctly!).

2.1 Data Scraping Task

You have to assigned a unique category for Daraz product as discussed in class. You can open your assigned category from the most left side of Daraz website or you can just search that category in search bar on the website. You have to scrap all products data (Product name, Product Price, Product Reviews) that is given on the webpage. You have to scrape minimum 4-5 pages data by automating the scraper to go to next page automatically.

2.2 Data Cleaning

After Scraping, you have to clean the data, by removing extra spaces, Nan, Webpages addresses, emails and invalid characters etc.

2.3 Perform Sentiment Analysis

You have provided two files. i.e. Positive, Negative. After Scrapping and cleaning you have to perform Sentiment analysis on that reviews by comparing positive and negative words from that file and assign the sentiment accordingly as discussed in class. In the end, you have to take average of all reviews for same product and assigning single sentiment.

2.4 Python Libraries for scraping

You can use beautiful soup, scrapy and selenium for scraping.