



SIMAD UNIVERSITY

GRADUATE STUDIES

Social Media Analytics - Master of Data Science

Course Assignment

This is the final project of the Social Media Analytics course and the **main objective** is to assess your knowledge by collecting your own social media data (from **any social media outlet**) and by creating and analyzing your own visualizations with the intention to discover interesting insights. **Remember**, you might need to (1) perform some data cleaning, (2) train a machine learning model and (3) reorganize the data in a format that works for the purpose of your analysis.

As far as this project is concerned, it is important that you collect your own dataset using Social Media Outlet API then, in a Python notebook, clean, train a model and create your own data visualizations. Once you have done that, the next step will be to explain your project by writing a report of 10 pages. The font format should be 12 point Times News Roman with 1.5 line spacing.

Note: It is prohibited to copy someone else's work whether it is a program code or a report. The assignment carries 60% of the total marks for this course.

Main tasks

For this part of the project, I created specific tasks I wanted students to answer (below).

- Task 1 - Let's gather our own data (e.g. minimum 1000 posts or records) related to a hot topic in Somali politics or sociology from a Social Media Outlet by using its API. Put into one CSV file for all the collected tweets that are in Somali language?
- Task 2 - Let's read the CSV file with some more analytical and statistical analysis like head(), describe() etc?
- Task 3 - Let's clean the data by removing the extra data and outliers, dropping or filling the missing values, etc. Create the final dataframe for further analysis?
- Task 4: Lets analyze and visualize the distribution of post length and word counts?
- Task 5 - Let's analyze and visualize the most trending posts?
- Task 6 - Let's analyze and visualize the top 5 users by number of posts?
- Task 7 - Let's visualize the prevalent words in posts using WordCloud?
- Task 8 - Let's visualize in what STATE/COUNTRY were the top posts posted that mention your TOPIC?
- Task 9 - Let's visualize on what day(s) of the month was your TOPIC talked about the most on the selected social media platform?
- Task 10: Let's collect 3,000 statements across Somali public pages on Social Media Outlet and annotate them into three sentiment labels:- Positive-wanaag, Negative-xumaan or Neutral-dhexdhexaad?
- Task 11 - Let's train a sentiment analysis model using machine learning?
- Task 12 - Let's now apply the trained model on the selected topic textual contents in order to predict their sentiments?

- Task 13 - Let's visualize the sentiment percentages (positive, negative, neutral) for the selected topic?
- Task 14 - Let's visualize the top positive keywords associated with the selected topic?
- Task 15 - Let's visualize the top negative keywords associated with the selected topic?

Presentation

Each student will present the results of his/her project in class and the presentation should be based on the final report. Each student will have five minutes for a power-point presentation.