Project Proposal for IAL621: Content Analysis for Social Network Data

Extract & Convert Social Network(Twitter, Facebook) content for sentiment analysis

Bala Mallampati

Master's in informatics and Analytics
University of North Carolina, Greensboro, USA
b_mallampat@uncg.edu

Problem Statement:

Understand the Social Network (Twitter, Facebook) and web scrap text information to do sentiment analysis of topics/tweets by following all privacy, regulatory & access policies.

In this project, we seek to answer the following data research questions:

A. Twitter network analysis to identify their networks as part of exploratory analytics.

B. Web scrap information from top tweets & topics from Twitter site and convert to numbers as part of Data wrangling.

C. Classify tweets/hashtags using Machine Learning Models

D. Do sentiment analysis of Tweets/Topics and present extracted insights to corresponding organization.

Introduction and Motivation:

Now a days each and everyone using Social Media for News, follow Trends and share their feelings and get in touch with their friends & family/Employers/Learning. Social media can create impacts on every organization if management is not up to date on latest trends/news.

I love to play with data, now "The world's most valuable resource is no longer oil, but data" as per The Economist article (https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data).

If we can understand how these social network works and can read & extract useful information then we can predict how it will impact our organization by doing sentiment analysis on Tweets/topics.

In future, I want to extend this project to use MassMine framework and integrate all social networks data include image & videos analysis to predict impacts on organization and take preventive actions.

Data Source and description:

Data will be extracted from below social media sites.

https://twitter.com/

https://www.facebook.com/

https://www.instagram.com/

Methodology:

Data processing is an important step for in the data analysis. Data science involves methods of analyzing massive amounts of data for the purposes of knowledge extraction. It evolved from statistics and traditional data management. Data comes in many shapes and forms, and many times we need to get it ready to be able to analyze it. The phrase "garbage-in and garbage-out" is particularly applicable to text mining to Train and Test Data.

Preliminary Plan:

References:
- Sklearn, Python library
- https://developer.twitter.com/en/portal
- https://towardsdatascience.com/3-super-simple-projects-to-learn-natural-language-processing-using-python-8ef74c757cd9