Sentiment Analysis of Corona in Lebanon

Manoli Elalam
School of Engineering
Lebanese American
Univesity
Byblos,Lebanon
manoli.elalam@lau.edu

***Abstract*—With the rapid development of online technology, information on the web is becoming vastly huge chunks of data that is hard to arrange and analyze. Due to this large amount of information, filtering this information and studying it is becoming more difficult with human resources. One of the advanced methods to research data is by studying its sentiment analysis, a new and famous area for research and social media analysis. In this paper we are analyzing the sentiment analysis of practically the topic of the year which is Covid-19. I will be focusing on 2 sentiments: positive and negative. Support by the twitter API, I've retrieved tweets of the Lebanese population, with the help of text mining tools, then studied their sentiments relating to corona. A trending research technique for a trending society topic.** **Keywords—Text Mining, Sentiment Analysis, NLP, RestApi, Opinion Mining, Covid19.**

I. INTRODUCTION

Sentiment analysis or opinion mining is the process of identifying and detecting subjective information using natural language processing. Its aim is to extract information on the attitude of the writer towards a specific topic. One of its main purposes is to conduct a quick, accurate and comprehensive analysis to study and address a certain issue, in a way in which human resources aren't able to do so. It's very efficient in filtering out valuable information in a short time. With the help of natural language processing and machine learning, this research method gives one of the best results in terms of precision and accuracy. Twitter was chosen as the platform for my database of information due to it being a legitimate communication channel about all social topics and its ease of use when it comes to information/tweets retrieval. The topic of interest in this paper is the corona virus and the its influence on the Lebanese citizens. In this paper I gave out the emotions and the sentiments of the Lebanese when it comes to Covid-19. By analyzing everyday tweets related to Covid-19 and showing its sentiment, I've saw it effects on our way of living, on our economy and our society.

II. BACKGROUND

Sentiment analysis started as a simple tool to analyze online product reviews then moving to social media texts from Twitter and Facebook, to now becoming a critical tool to have in every company. What made opinion analysis a trending research topic was the possibility to automatically collect and analyze large corpuses of opinions with the help of text mining tools. A fast and simple method for determining the sentiment of a text is using a pre-defined collection of sentiment-bearing words and simply aggregating the sentiments found. I used python as my main programming language because of its ease of use and the natural language processing libraries found in it. Moreover, its support for the Twitter API was the key to this research. Twitter offers two APIs to retrieve data from tweets: REST and Streaming. For this paper, the REST API was used. The tweets were retrieved in form of the JavaScript Object Notation (JSON) documents. After cleaning the tweets from noisy words, the tweets were stored along with their sentiments in a data file (.txt). For getting sentiments, text blob's method was used to classify the tweets as positive or negative. In order to demonstrate the effects of the Covid19 pandemic, this research will identify local tweets being currently tweeted, define appropriate tweets whether related or not to our topic, determine its main point, develop an analysis about everything corona and demonstrate on what basis it is positive or negative sentiment.

III. IMPLEMENTATION

- The 1st step was getting the twitter API keys from the twitter development console. After acquiring them I had to establish a connection between the program and twitter through authentication. From there we could fetch any information from twitter.
- Using a query to fetch the target word in JSON format then printing out just the tweet
- Get tweet sentiment through a utility function that classifies sentiments of passed tweets using Text Blob's sentiment methods.
- Clean tweets and filter them through various methods discussed later
- Classify them as positive/negative then analyze

IV. Discussion

A. Data Cleaning

Twitter gives its user the ability to search a specific key word. It was essential to get all corona updates. In my case I looked for 'corona','covid19','pandemic' tweets. It was all done through a query from the twitter development console. Moreover, from the get_tweets function, I filtered retweets to avoid getting retweets hence unwanted duplicates. Setting the language to English guaranteed that we would retrieve English based tweets that we could understand and analyze.

In order to get the most accurate results and disregard unwanted bait tweets we sat out for cleaning the tweet. As a 1st step we wanted to remove unnecessary symbols. A utility function to clean tweets by removing link, special symbols was used using simple regex statements. Still though, we noticed that the data contained a lot of stop words and some slangs that was needed to be replaced or even removed completely. Removing stop words was done by importing from the Nltk.corpus library a certain array of stop words, then filtering the tweet, if it had words similar inside the array of stop words, we removed them from the tweet. Moreover, slang words and abbreviations were a problem that needed also justifications. This was done by applying a public function called translator.

B. Showing Sentiment

I was interested in getting the sentiments from Lebanese citizens and practically data related to Lebanon. I added a Geofilter location that targets tweets coming out from just Lebanon to disregard unwanted data.

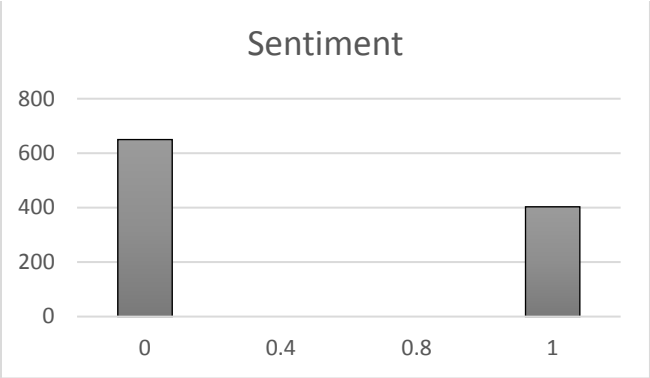
By using text blob's sentiment method, the machine learning part and studying sentiments were taken care off due to this library. It uses a polarity system, such that if the score of a tweet was >0 it is set to positive, =0 means neutral and <0 returns negative sentiments. I created 2 arrays (ptweets and ntweets) to store positive and negative tweets respectively. If its sentiment is positive, the tweet shall be stored in ptweets and similarly ntweets. Stored so it can be later printed under its category. The percentage was also calculated to show final result and what was the outcome of the Lebanese

V. Results

Tweets were gathered for 2 weeks gathering approximately 1000 tweets in that period and stored in a data file (.txt). Each day had its own sentiment of positive/negative percentages. A sample output of showing sentiment had the form of:

Positive tweets percentage:26.666666666666668 %
Negative tweets percentage: 46.666666666666667 %
Positive Tweets: ----------
Negative Tweets: ----------

After combining the data into 1 file, we ran the code without using public tweets but by reading the tweets from the text file.



The following results where obtained where 1 shows a positive sentiment and 0 a negative sentiment. It is evident that the negative results outweighed its positive. Expectedly corona is leaving a negative impact on Lebanon.

	Percentage	Number
Positive	38.2716%	403
Negative	61.7283%	650

Over the 2 weeks, the Lebanese citizens expressed negative thoughts of around 61% while just 38% positive expressions.

Some notable topics that aided in increasing the negative percentage were found in parallel with corona were the revolution, the economic crisis, the lack of jobs and funds from the government to help the needy. Moreover, thoughts about staying in quarantine really made the Lebanese unhappy showing how much they like to go out and have fun.

On the other hand, the majority of the positive tweets came from how the corona situation was handled and how the curve is barely increasing. The red cross of Lebanon took great recognition from its citizens. The red cross also was one of the top tweeters providing daily updates to corona cases. Also, positive tweets showed up about how the Lebanese citizens are helping each other in this tough time.

VI. Conclusion

As a conclusion, it was obvious the Covid19 Pandemic greatly affected the Lebanese positively and negatively. Most of the positive data were about how the pandemic was handled inside in the country, the red cross and how the people came together to help each other during this tough phase. Majority of the negative tweets were of course the health issue, the bad economic state we are in because of the lack of work and an uprising revolution. Moreover, negative tweets about being in quarantine showed how much the Lebanese like to live their life and be happy.
In the future the application areas of sentiment analysis will increase, and it will be a critical part in analyzing huge chunks of data for many different purposes. Moreover, we will see that the research methods involving sentiment analysis will improve due to advances in natural language processing and machine learning opening up unlimited research opportunities.

Acknowledgment

This research has taken a considerable amount of time and resources and I would like to acknowledge the help of Dr. Haidar Harmanani for helping me and guiding me throughout this journey to present this research. His expertise in the data mining field was critical and beneficial for the core of this project. Thank you for giving me this opportunity to try and explore something new in the computer research field. I would also like to thank the Lebanese American University for always encouraging us to reach higher heights in our careers.

References

[1] "Naïve Bayes Classifier"-May 10,2010 JacobJ.

[2] "Research on Sentiment Analysis of Financial Texts Based on Web". College of Mathematics and Information Science

[3] The Evolution of Sentiment Analysis - A Review of Research Topics. Mika V. Mäntylä Daniel Graziotin, Miikka Kuutila .Finland 2013

[4] Analyzing Political Sentiment on Twitter. Martin Ringsquandl and Dušan Petković.University of Applied Sciences Rosenheim. 2013 AAAI Spring Symposium

[5] Research On Sentiment Analysis:The First Decade. 2016 IEEE 16th International Conference on Data Mining Workshops. Oskar Ahlgren.

[6] Basic Data Cleaning/Engineering Session] Twitter Sentiment Data . Jae Duk Seo.May 28,2018. https://towardsdatascience.com/

[7] How to Categorize Large Amounts of Text Data. Data@Urban Feb 29,2019. *https://medium.com/*