ABSTRACT

KEYWORDS – Twitter, Sentiment Analysis, Weather data

INTRODUCTION

Sentimental mining is a critical and essential area because sentiment fundamentally relates to a person’s emotions, impression and attitude. Analysing every individual’s sentiments from a text analytics point of view accurately is challenging. We are basically trying to comprehend the thoughts and assumptions of an individual regarding a concept or topic with some context be it positive, negative or neutral [1]. It is estimated that nearly 2.5 quintillion bytes of data is generated each day [2]and there is a plethora of information relating to a person’s messages, tweets, documents, emails, chats, conversations and comments available. Sentiment analysis aids us in analysing this huge amount of data efficiently.

In this paper, we focus on performing sentiment analysis on the weather based on data retrieved from Twitter. [3] Twitter, being an acclaimed stage for social networking possess and people to express their interests and thoughts has plenty of information from posts known as “tweets”. With over 500 million tweets per day pertaining to almost anything, it is an excellent platform due to its accessibility and real-time analysis of the data which is crucial for sentiment analysis. It is not uncommon that weather plays an important role in an individual’s mood and its consequences because of the decisions made. **[I need some help in writing what exactly we plan to do with the weather data… for instance determine if the person is happy or not by performing sentiment analysis on tweets regarding weather? Basically, we need to write something that sets us apart from all the previous work done using these 3 even if its simple (sentiment analysis, twitter data and weather)]**

In its preliminary stage, we [need to write what we have done so far]

RELATED WORK

As the accessibility to real-time information or otherwise is increasing, performing analysis on this data is also becoming viable. There has been a lot of notable research done on sentiment analysis of twitter data in general and studies emphasizing on [4] correlating weather with people’s mood and sentiments from Twitter data by measuring temperature, humidity and atmospheric pressure. These weather variables seem to have a satisfactory impact on one’s mood. [5] This paper focuses on implementing sentiment analysis on Twitter data using the Twitter API’s and a wealth of available libraries. Twitter is known for having small texts and abbreviations used by people in their posts or tweets making it challenging to extract polarity of the texts and hence researchers resort to utilizing deep learning and machine learning techniques.

Researchers have performed sentiment analysis for various reasons and different areas like elections, politics, movie ratings and fashion to name a few. [6] In 2015, there was a noteworthy research done in the vision of predicting future crime on each area of a major city, Chicago, Illinois of the United States using GPS tagged twitter data. They aimed to predict the time and location during which a specific type of crime is expected to occur by applying lexicon-based sentiment analysis on categorized weather data combined with kernel density estimation of historical crime incidents and were successful. [7] This paper concentrates on using a Twitter specific sentiment extraction methodology and explore a corpus of over 1.5 billion tweets. With the help of machine learning techniques on Twitter corpus correlated with the weather at a particular time and location of the tweets, it was concluded that aggregate sentiment follows different climate and seasonal patterns.

METHODOLOGY

Twitter Data

It is estimated that nearly 2.5 quintillion bytes of data is generated each day [2]and there is a plethora of information relating to a person’s messages, tweets, documents, emails, chats, conversations and comments available. Twitter, being an acclaimed stage for social networking possess and people to express their interests and thoughts has plenty of information from posts known as “tweets”. With over 500 million tweets per day pertaining to almost anything, it is an excellent platform due to its accessibility and real-time analysis of the data which is crucial for sentiment analysis. It is essential to figure out the sentiment of each tweet whether it is positive, negative or neutral. We utilised [8] Archive Team: Twitters’ Stream Grab of the year 2018 which is a simple collection of JSON grabbed from the general twitter stream. **[Our dataset consists of XXX number of tweets and XXX users. – This can be removed if you guys think it is not required]**

Weather data

The historical weather data was collected from Met Eireann, the Irish National Meteorological Service which is the leading provider of weather and information related services for Ireland [9]. [10] Wayback-machine-scraper is a repository which consists of command line utility that can be used to scrape data. We utilised it to download all the saved snapshots from the Met Eireann website for the year of 2018. [11] Boards.ie is a discussion board with a wide range of forums which includes weather. A registered users’ daily forecasts weather forum was utilised as another additional source of data by exporting the print versions of pages spanning the year 2018.

RESULTS  
DISCUSSION and CONCLUSIONS

NOTES  
1 – We need to add some images, charts or some sort of visualization in the paper with the results obtained as of now. The links that Cian has provided might be of some use (listing them here again)

<https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/>

<https://www.softwareadvice.com/resources/free-twitter-sentiment-analysis-tools/>

2 – **[Separate program to extract tweets based on location fields from tarball and serialize to CSV format (twittermunge)], Also polling** [**http://www.met.ie/opendata**](http://www.met.ie/opendata) **for new forecasts** - Cian could you please add lines into the essay regarding this accordingly.

REFERENCES

1 - <https://blog.aylien.com/why-is-sentiment-analysis-important-from-a-business-perspective/>

2 - <https://blazon.online/data-marketing/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>

3 - <https://towardsdatascience.com/mining-twitter-data-ba4e44e6aecc>

4 - <https://pdfs.semanticscholar.org/b282/feb759e57530b115dcc4bb080f96598a5246.pdf>

5 - <https://arxiv.org/ftp/arxiv/papers/1711/1711.10377.pdf>

6 - <https://www.researchgate.net/profile/Xinyu_Chen20/publication/282380303_Crime_prediction_using_Twitter_sentiment_and_weather/links/5ac78afda6fdcc8bfc7fa59c/Crime-prediction-using-Twitter-sentiment-and-weather.pdf>

7 - <https://www.ccs.neu.edu/home/amislove/publications/Weather-ICWSM.pdf>

8 - <https://archive.org/details/twitterstream?and%5B%5D=year%3A%222018%22&sort=titleSorter>

9 - <https://www.met.ie/>

10 - <https://github.com/sangaline/wayback-machine-scraper>

11 - [https://www.boards.ie/vbulletin//showthread.php?t=2055579971](https://www.boards.ie/vbulletin/showthread.php?t=2055579971)