Twitter Sentiment Analysis

*Abstract*—Social network is a technology globally accessible by millions of people, resulting it to be a fundamental platform to generate ideas, reviews and opinions. Social media users tend to share information including their location in respective profiles. Thus, this Social network crawl will act as a tool used to obtain required data via several processes, which is in line with specific procedure and policy, and researchers will be able to get real time list of trending topics detected by this method. Social network crawl plays crucial role as a first step before data analytics and visualization. Crawl is a process of extracting unformed data from social network sites through various methods and separate the unwanted elements to produce formatted required data set according to keywords search and location. The data will be programmatically collected from social networks through their API and detect top terms in the data. Data collection will be covering important locations or cities in Malaysia. Specific approach made to identify trending topics is by ranking the terms according to spikes in terms frequency. Therefore, using this framework, researches/users will be able to detect topic trends, audience engagement, brand monitoring, sentiment analysis and many more for a given time period and location according to social network collected data

Keywords—sentiment analysis, user research, twitter, natural language processing.

# Introduction

Sentiment is a program that rate the sentences, which tell the sentence in positive, negative or neutral. A crawler is a program which visits web pages and read entire data in accordance to create record for search engine index. This ccrawler also sometimes called as a “spider”, “internetbot” or “spiderbot” and often shortened to crawler. It is programmed in a way to searching for something specific in the World Wide Web and cater up-to-date data to the search engines. Crawler can manage a large set of data. The number of Internet pages is tremendously large; even the largest crawlers fall short of making a complete index. For this reason, in the early years of the World Wide Web before 2000, search engines had difficulties to give specific or filtered relevant search results. According to the application and user group, it may be advantageous to preferentially acquire pages that are highly linked, pages that relevant to specific topics, pages that are likely to mention specific structured information, pages that score highly with respect to queries submitted frequently to the search engine, pages that change frequently, and so on. Focused Web. The Twitter Sentiment Analysis will enter the keyword that want to search, after that it will crawl the data from twitter who ever tweets and analyse the data, which that sentence is positive, negative or neutral.

# Sentiment analysis

The area of study that interprets people’s opinions, against any particular topic, about any event etc. in text mining it known as opinion mining or sentiment analysis. It produces a vast problem zone. There are also various names and having different tasks, e.g., sentiment analysis, opinion extraction, opinion mining, sentiment mining, affect analysis, subjectivity analysis, review mining, etc (Liu Bing, 2012).

## Document Level Analysis:

This level classifies that whether the complete document gives a positive sentiment or negative sentiment. The document is on single topic is considered. Thus, texts, which comprise comparative learning, cannot considered.

## Sentence Level Analysis:

The task of this level is sentence by sentence and decides if each sentence represents opinion into negative, positive, or neutral. Neutral, if sentence does not give any opinion means it is neutral. Sentence level analysis related to subjectivity classification. That expresses information from sentences that gives subjective aspect and opinions. I.e. good-bad terms.

## Entity/ Aspect Level Analysis:

Both the document and the sentence level analysis do not find peoples like and dislikes. Entity/Aspect level gives throughout analysis. Entity/Aspect level earlier called feature level. The core task of entity level is to identification constructs, aspect level straight forward give attention at the opinion or sentiment. It is based on the concept that an opinion resides of an attitude and a destination of opinion.

# Related Work

There are various text mining approaches used to mine the

data.

Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government., (Ravinder, 2017) have examined this government policy the demonetization from the ordinary person’s viewpoint with the use of the approach of sentiment analysis and using Twitters data, Tweets are collected using certain hashtag (#demonetization). Analysis based on geo-location (State wise tweets are collected). The sentiment analysis API used from meaning cloud and classified the states into six categories, they are happy, sad, very sad, very happy, neutral, and no data.

Sentiment Analysis using product review data (Fang, Xing, 2015) have solved the issue of sentiment polarity categorization, and it is one of the basic problems of sentiment analysis. Online product reviews data used in this study, collected from Amazon.com. In this paper, Investigation for both sentence-level categorization and review-level categorization achieved. Scikit-learn software used for this study. Scikit-learn is an open source Machine Learning software package in Python. Naïve Bayesian, Random Forest, and SVM: These classification techniques selected for categorization.

Sentiment analysis of twitter data using Machine Learning approaches and semantic analysis (Divakar, 2014) contribute to the sentiment analysis for customers’ review classification. Already labeled twitters data used in this task. They have used three supervised techniques in this paper: naïve-Bayes, Max-entropy and SVM followed by the semantic analysis that was use along with all three methods to calculate the similarity. They have used Python and NLTK to train and classify the naïve-Bayes, Max-entropy and SVM. Naïve-Byes approach gives a better result than the Max-entropy and SVM with unigram model gives a better result than using SVM alone. Then the correctness then increased when the Word-Net of semantic analysis applied after the above procedure.

Sentiment analysis in twitter using Machine Learning techniques (Neethu, 2013) in this paper, they analyze the twitter data related to Electronic products using Machine Learning approach. They existent a new Feature-Vector for classification of the tweets and extricate peoples’ opinion about Electronic products. Thus, Feature-Vector is created from 8 relevant features. The 8 features used are special keyword, presence of negation, pos tag, and number of positive keywords, emoticon, and number of negative keywords, number of negative hash tags and number of positive hash tags. Naïve-Bayes and SVM classifiers are implemented using built in functions of Matlab. Max-Entropy classifier is implemented using Maximum-Entropy software. All the used classifiers have almost equal performance.

# Twitter

The aim while performing twitter sentiment analysis is classifies the tweets in different sentiment classes accurately. In this field of research, various techniques have evolved, which come up with methods to train a model and then test it to check its effectiveness. Performing sentiment analysis is challenging on twitter tweets. Here we define some reasons for this:

•Limited tweet size: with just 280 characters in hand, reduced statements are generate, which results sparse set of features.

•Use of slang: these words are not quite the same as English words and it can make an approach out dated because of the evolutionary use of slangs.

•Twitter features: it permits the use of hash tags, user reference and URLs. These require different processing in comparison to other words.

•User variety: the users express their opinions in a different ways, some using different language in between, while others using repeated words or symbols to express their emotion.

All the above problems required to face in the pre-processing section.

# Sentiment Analysis on Twitter Data

The workflow for sentiment analysis shown in Figure 2. The system consists of the four main modules: data collection module, data processing module, classification module and analysis of output.



Figure : Twitter Sentiment Analysis Workflow

A. Input (Keyword): First start by choosing a subject then we will collect the tweets with that keyword and perform sentiment analysis on those tweets.

B. Tweets Retrieval: Tweets can be a structured, semi-structured and unstructured type. Sentiment Analysis research, we can collect tweets using different programming languages like R or python.

C. Pre-Processing Data pre-processing is nothing but filtering the data to remove the incomplete noisy and inconsistent data. Following tasks are involved in pre-processing task:

• Removal of Retweets (in case of twitter dataset)

• Removing URLs, Special characters, Punctuations, Numbers etc.

• Removing Stop words

• Stemming

• Tokenization

D. Sentiment Detection: Sentiment word identification is important work in many applications of sentiment analysis and opinion mining, such as tweets mining, opinion holder finding, and tweet classification. Sentiment words can be classify into Positive, Negative and Neutral words. The fundamental task in Sentiment Analysis is classifying the polarity of a given tweets feature. The polarity is in three classes i.e. Positive, Negative and Neutral. Polarity identification done by using different lexicons e.g. Bing Lui sentiment lexicon, SentiWordNet etc. that help to calculate sentiment strength, sentiment score, etc.

e. Classification Algorithm:

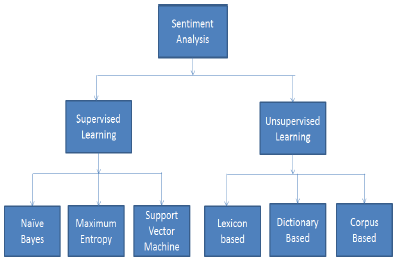


Figure Sentiment Analysis Algorithms Table Type Styles

Two fundamental approaches are there in sentiment analysis i.e. Supervised learning Approach and unsupervised learning Approach. Sentiment classification of twitter data done using supervised machine learning approaches like Naïve-Bayes, SVM, and Maximum-Entropy etc. Efficiency of classifier built upon which dataset used for which classification methods. In the case of Supervised machine learning approaches to train the classification model Training dataset is used which then help for classification of test data.

F. Analysis of Output: The fundamental thought of sentiment analysis is to change unstructured data into the significant or meaningful data. After the completion of analysis, the results displayed on graph like pie chart.

# Conclusion

The analysis of Twitter data been done in different points of view to mine the opinion or sentiment. This program defined the concept of sentiment analysis and opinion mining with respect to various levels of sentiment analysis. This survey paper discussed different techniques of sentiment analysis and methodology for sentiment analysis. If we are doing twitter sentiment analysis, it is necessary to know about the twitter, about extricating the tweets, its structure, and their significance. This paper gives brief idea of tweets. Thus, the essential information required to do sentiment analysis of Twitter well discussed in this review paper.

##### References

1. Surnar, Avinash, and Sunil Sonawane. "Review for Twitter

Sentiment Analysis Using Various Methods.".

1. Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5.1 (2012).
2. Singh, Prabhsimran, Ravinder Singh Sawhney, and Karanjeet Singh Kahlon. "Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government." ICT Express (2017).
3. Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Contemporary computing (IC3), 2014
4. Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. IEEE, 2013.