



**FEU Institute of Technology**  
COLLEGE OF ENGINEERING • COLLEGE OF COMPUTER STUDIES

**ITBSPEC5:  
ANALYTICS APPLICATION**

Professor: Maria Vicky S. Solomo

**Twitter Sentiment Analysis using  
Orange Data Mining Tool**



Submitted by:

AGARAP, CHRIS JOHN  
BALDONADO, REX CHRISTIAN  
CALIMLIM, VEO THADEUS  
ROLLAN, KERVIN

# *Twitter Sentiment Analysis using Orange Data Mining Tool*

Agarap, Chris John M.

ITE Department – ITSMBA

FEU Institute of Technology

SMDC Sun Residences, Mayon St., Quezon City

chrisjohn.seej@gmail.com

Baldonado, Rex Christian B.

ITE Department – ITSMBA

FEU Institute of Technology

1856 Road 13, Fabie Estate, Paco, Manila

rcb122398@gmail.com

Calimlim, Veo Thadeus V.

ITE Department – ITSMBA

FEU Institute of Technology

70 General Ave., Tandang Sora, Quezon City

veothadeus@hotmail.com

Rollan, Kervin P.

ITE Department – ITSMBA

FEU Institute of Technology

Natividad Lopez St., Ermita, Manila

kervin\_rollan@yahoo.com

**Abstract**—Since social media nowadays has become an integral part of life, the rise of the different social networking platforms is inevitable. One example is the Twitter which is a free social networking service that allow people to share their thoughts and opinions called tweets. With millions of people tweeting every day, the researchers became curious as to what are the emotions behind those tweets. The researchers have used the software Orange to classify the tweets based on Ekman's 6 Universal Emotions Theory and identify whether the tweets are positive, negative, or neutral. 1000 tweets were gathered using the Twitter API via its query search for content. Before using the fetched data for analysis, the tweets were preprocessed using the preprocess text widget to remove unnecessary words or punctuation marks commonly found in the word cloud created beforehand. Sentiment Analysis was used to determine and predict the emotions behind the tweets of each user. The researchers used Vader's technique in sentiment analysis widget and Ekman's emotion classifier in tweet profiler widget. To visualize the results of the analysis, boxplot, distribution chart, and heat map were used.

**Keywords**—Sentiment Analysis, Twitter API, Orange, Tweet Sentiment, Identify Emotions

## I. INTRODUCTION

Social media is becoming an integral part of life online as social websites and applications continue to rise. One of the most popular examples of social media is Twitter. Twitter is a free social networking blogging service that allows registered members all over the world to broadcast short posts called tweets. It was established on March 21, 2006, by its founder Jack Dorsey. With millions of tweets being posted by minute you have to wonder, what are the stories and emotions behind it. That is why the researchers have chosen this topic to solve the curiosity.

It is nearly an impossible task considering the many different factors to consider when analyzing tweets. Are those tweets supposed to mean that way? Are those tweets directed to someone? And so much more. That is why the researchers incorporated the software Orange to be able to utilize its built-in features in conducting sentiment analysis to the tweets.

## II. BACKGROUND OF THE STUDY

One of the best things about Twitter is perhaps its greatest appeal is in its accessibility. It's easy to use both for sharing information and for collecting it. But Twitter and all that accessible information have some undeniable downsides, too. For one thing, anyone can say just about anything.

Dr. Drexler P. (2014) stated that via Twitter, the ever so tempting of attacking someone who has wronged you before or being attacked is always a possibility igniting a sort of "Twitter war" we often see.

One of the methods that will help in the analyzation of tweets is the sentiment analysis wherein it identifies and categorizes opinions expressed in a text. Sela R. (2015) claimed that with the utilization of sentiment analysis on Twitter, people can monitor then figure out how the user reacts to a particular person, brand or event.

The Twitter Sentiment Analysis using Orange Data Mining aims to fulfill the following:

- To identify the emotion of tweets from the Twitter users based on the Ekman's 6 universal emotions theory.
- To identify whether the tweets are positive, negative or, neutral.

### III. METHODOLOGY

The data set came directly from Twitter using the Twitter API. To fetch tweets using the API, credentials are required – consumer key and consumer secret – which are obtained in [apps.twitter.com](https://apps.twitter.com). The researchers have fetched 1000 tweets via its content for a query search. Some keywords that are used for searching tweets are sad, happy, disgust, anxious, fear, and annoyed. These words are some of the possible ways of discovering emotions in tweets.

Twitter (1) ? X

### Twitter API Key

Query

Query word list: sad happy depressed suicide anxious

Search by: Content

Language: Any

Max tweets: ☒ 1000

Allow retweets: ☐

Collect results: ☒

Text includes

☒ Content ☐ Author Description

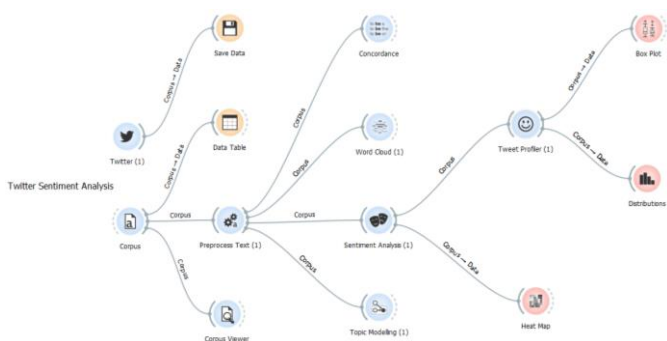
Info

Tweets on output: 1000

Report Search

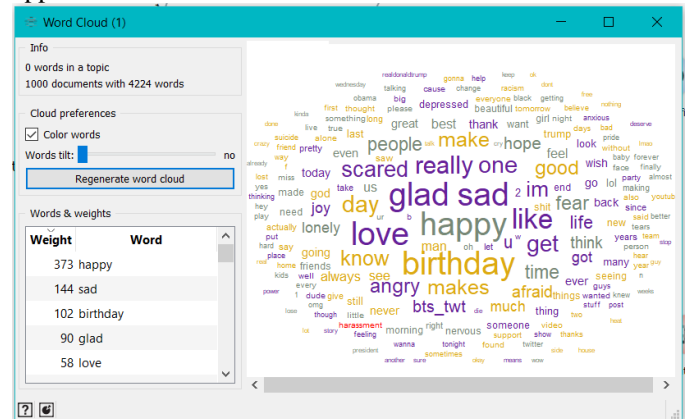
? [Icon]

In this study, the researchers used Orange Data Mining – a simple and interactive tool wherein data analytics could be learned without the need for programming experience. The add-on Text Analytics was installed on the application in order to utilize some APIs and text analytics techniques such as the Sentiment Analysis that is used in the study. Sentiment Analysis is the process wherein texts are mined to identify and extract information (Gupta, 2016). This analysis is used in the study by identifying or predicting the emotions contained in a tweet.




## IV. APPLICATION DESIGN AND IMPLEMENTATION

Since the tweets gathered contain unnecessary character or words, they were preprocessed first to clean the data. The result of the preprocessed data could be visualized using the word cloud wherein the frequency of words in the tweets are easily seen with the help of their sizes – how big or small the words appear.



Once preprocessed, the data is now ready for the sentiment analysis. In Orange, the sentiment analysis contains two techniques which are Liu Hu and Vader. The two differ in terms of how the score of sentiment is computed. Liu Hu computes for a single normalized score while Vader computes for the score in each category. This means that Liu Hu is a simple approach that does not require the data to be trained while Vader requires it because of its complexity in computing the score per category.



To view the result of the sentiment analysis, the tweet profiler is used. Using this widget, the category of the emotion of the tweets are seen using different emotion classifiers. These classifiers include Ekman, Plutchik, and POMS. The researchers used Ekman that categorizes the tweet emotion in terms of anger, disgust, fear, joy, sadness, and surprise.

Tweet Profiler (1)

Options

Attribute: **S** Content

Emotions: Ekman (Multi Class)

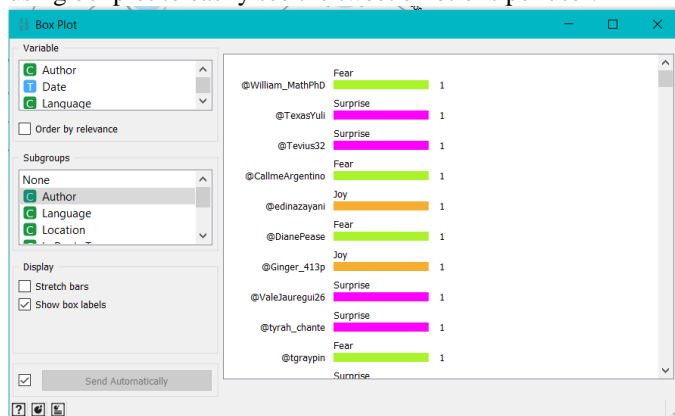
Output:

Report ☒

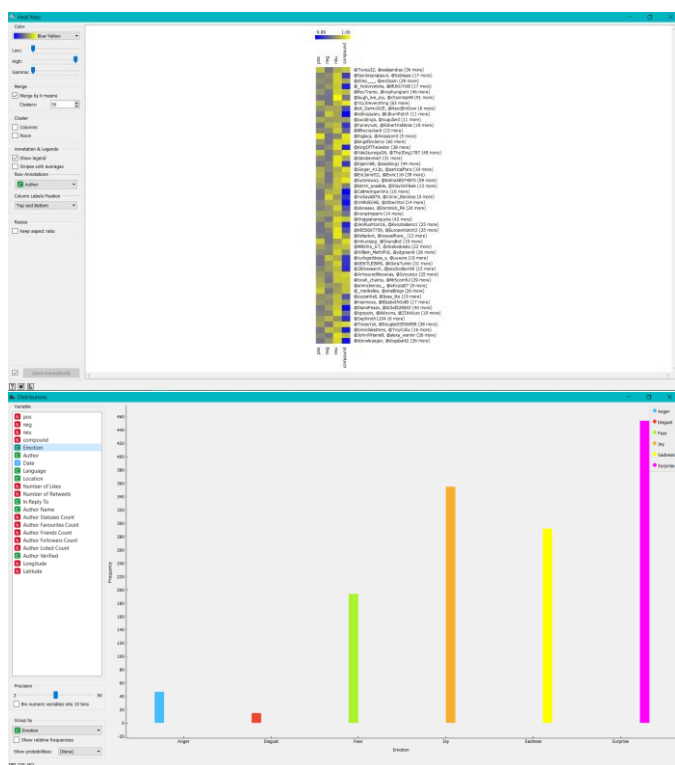
? [File Icon]

- Ekman (Multi Class)
- Ekman (Multi Label)
- Plutchik (Multi Class)
- POMS (Multi Class)
- Ekman (Multi Label)
- Plutchik (Multi Label)
- POMS (Multi Label)

Lastly, the result gathered from the analysis is visualized using boxplot to easily see the tweet emotions per user.



Another visualization used are the heat map to easily differentiate the values of the analysis based on colors and the distribution to view the frequency of the emotions.



## V. DISCUSSION

Based on the Sentiment Analysis that was implemented, some Twitter users have multiple emotions based on how many

tweets are fetched. Majority of the users have resulted in only one emotion because only one tweet was fetched using the Twitter API. When multiple tweets are obtained per user, the tweet profiler shows multiple emotion based on the content of the tweet generated by the sentiment analysis.

The emotions produced are based on Ekman's Six Universal Emotions Theory – anger, disgust, fear, joy, sadness, and surprise.

This study could be used not only professionally but also personally. A company could use the tweets of users to understand the social sentiment of their brand, product, or service and to gain insights on how these users feel about them. Personally, an individual could use the study by understanding the feelings of a friend or family member based on their tweets.

## VI. RECOMMENDATION AND FUTURE WORKS

The researchers recommend another theory of emotions besides the Ekman's theory.

The group also recommend that the future researchers should consider the retweets because of the possibility of the user agreeing to the tweet. However, one downside of this is the duplication of tweets.

The result of the sentiment analysis will be improved if the tweets are fetched based on the author, not only from the content.

## VII. REFERENCES

- Dr. Drexler, P. (2014, February 21). *Twitter: The Good, the Bad and the Ugly*. Retrieved from Hello Giggles: <https://hellogiggles.com/lifestyle/twitter-good-bad-ugly/>
- Gupta, S. (2016, January 8). *Sentiment Analysis: Concept, Analysis and Applications*. Retrieved from Towards Data Science: <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>
- Sela, R. (2015, January 18). *What You Need To Know About Social Media Sentiment Analysis*. Retrieved from Curatti - The Editor of Chaos: <https://curatti.com/social-media-sentiment-analysis/>