Clustering Indonesian Cyberbullying Words in Social Networks

Hendro. Margono, Xun. Yi, and Gitesh. Raikundalia

***Abstract*—**Social media is a popular medium to communicate with each other through cyberspace. As a media for the interaction between users, social media is very exposed to the access of bullying innocent victims. The acts of harassing or bullying someone, typically by sending offensive and personal messages through online is called cyber bullying.

Mining Indonesian cyber bullying words in Twitter is important information for society. The first author has captured and downloaded tweet, which contains the Indonesian bullying words in Twitter, through Twitter Adder and Twitter Application Programming Interface (API). This study mines the Indonesian bullying words in Twitter by using K-mean clustering. K-mean clustering is successful in uncovering eight clusters that reflect the feature or group of Indonesian bullying words. This research aims to identify cyber bullying messages that has similar contents for the intentions to bully into several groups. Every cluster contains some Indonesian bullying words which have a similar characteristic in their cluster. Moreover, the result of this study contributes important data about cyber bullying in Twitter. The finding results will provide important information for the public to be more attentive of bullying through social media.

***Keywords*—** Social computing, data mining, cyber bullying.

I. Introduction

IN general, sending messages can be broadly categorized into text and picture. Text is more expressed in an objective manner, so they mostly contain the intended words of something written in media such as paper and electronic devices. Generally, images are more emphasizing the ideas, but this works only focuses on Twitter text messages that contain Indonesian bullying words.

Bullying on the Internet is commonly popular for adolescents in Indonesia. The National Child Protection Commission (KOMNAS ANAK) has conducted a survey to identify the proportion of children in Indonesia receive abuse from their peers. KOMNAS ANAK had reported that 87.6 percent of 1,026 respondents said that they suffered from mental, physical or verbal abuse and 42.1 percent of respondents who reported ill treatment proclaimed that they were bullied by their class-mates [1]. Akamai reported that Indonesia is the place of the worst cyber bullying occurred compared to China and The United States [2]. The percentage of cyber bullying in Indonesia has increased significantly from 21% to 38%. This increase illustrates that there is less attention from parents, government and NGOs to guide the children about bullying on the Internet [2].

Regarding the fact that the phenomenon of cyber bullying cases in the social media has increased significantly, analysing Indonesian bullying words in Twitter using clustering is an interesting topic, because there are some Indonesian expletives associated with bullying such as a bangsat (rascal), anjing (dog), monyet (monkey), etc [3]. These words are very impolite and insulting when they are being spoken to Indonesians, so it is considered as expletive words. These words illustrate the perpetrators' expression when they feel irritated with the victim by using bullying words associated with animals, psychology, disability, and attitude.

This work uses a clustering method to perform analysis of existing Indonesian bullying words. The clustering methods will explore the connection between Indonesian bullying words to make an assessment of their structure based on the similarity and the distance between Indonesian bullying words. Cluster analysis is applied to discover the natural groups of a set of patterns, points, or objects [4].

This research uses K-mean clustering. K-mean clustering is a method to partition *n* observations about Indonesian bullying words into K clusters in which each observation belongs to the cluster with the nearest mean [5]. The purpose of using K-mean clustering is to characterize Indonesian bullying words in similar groups. The operational K-mean in this work is as follows: Given *n* observation of Indonesian bullying words, this work variable to find K groups based on a measure of the distance and the similarity between Indonesian bullying words in the same group.

To run the clustering algorithm multi-dimensionally, this work uses Rapid Miner software in order to analyse Indonesian bullying words with K-mean clustering. Some processes are needed before analysing using K-mean method: first, importing data from the repository in Rapid Miner; second, filtering words to clean up unstructured sentences; third, using K-mean clustering techniques to measure the distance and similarities between Indonesian bullying words

Hendro. Margono is a postgraduate student in College of Engineering and Science. Victoria University. PO Box 14428, Melbourne 8001, Victoria. Australia. (e-mail: hendro.margono@live.vu.edu.au)

Xun. Yi, College of Engineering and Science. Victoria University. PO Box 14428, Melbourne 8001, Victoria. Australia. (e-mail: Xun.yi@rmit.edu.au)

Gitesh. Raikundalia, College of Engineering and Science. Victoria University. PO Box 14428, Melbourne 8001, Victoria. Australia. (e-mail: Gitesh.Raikundalia@vu.edu.au)

in the same group.

The result of this work is an important contribution to society for understanding insulting words associated with the Indonesian bullying words which have occurred in Twitter. This result also contributes in guiding Indonesian teenagers with cyber bullying in social networks.

This paper is organized as follows. Section 2 describes the related work of this paper. In section 3, we describe K-mean clustering which will be used to analyse the research problem. In section 4, the implementation of K-mean clustering to improve this research will be detailed. Analysing the similarities and the distance among words will discover new Indonesian bullying words patterns. The last section is the Conclusion section.

II. Related Work

Some similar topics have been discussed in previous research. Research work which used K-mean clustering is to find linking between words in microblogs and Wikipedia. Xu and Oard [6] used a clustering method to find the linking terms in microblog posts to Wikipedia pages and then to leverage Wikipedia's link structure in order to estimate semantic similarity. The results of the research show that they found the linking terms in microblogs of Twitter and Wikipedia pages.

Rosa, et al. [7] has also performed a research based on language similarity in social media by using hash tags in clustering approach. The techniques have been successful in finding the classification tweets based on language similarity rather than topical coherence.

Weng and Lee [8] have conducted research to detect events in Twitter. The technique that has been used to identify events in Twitter is Event Detection with Clustering of Wave-let-based Signals (EDCoW). EDCoW has been successful in analysing individual words in Twitter by applying wavelet analyses on the frequency based on raw signals of the words. Then the words are clustered to form events with a modularity-based graph partition technique.

Analysing context words in social media such as Twitter using the machine learning technique has been covered in previous research. Bann [9] constructed Latent Semantic Clustering (LSC) to evaluate the distinct language that had been used in tweets as an expression of certain emotion and perception. Bann's research demonstrated that the LSC method can help to detect visual emotion in Twitter. Moreover, his research is also successful at classifying tweets triggered by emotion and perception [9].

Recently, Chen and Zhao [10] conducted research about Twititude to analyse comments in Twitter based on clustering messages. Their research had developed Twititude to analyse content, summarization and event extraction in Twitter. The research had also developed document clustering by using some algorithms such as Latent Semantic Analysis (LSA), Non-negative Matrix Factorization (NMF), Latent Dirichlet Allocation (LDA) and Gamma-Poisson (GaP). Moreover, Chen and Zhao [10] stated that their research was successful in extracting automatically different aspects of talking about a subject, summaries of people's attitudes and comments towards each aspect.

Another research work which used the clustering method is Vector Space Model using Term Frequency (TF)-Inverse Document Frequency (IDF) weighting score [11]. The research develops a framework to analyse Twitter messages. The result of this research is discovering groups of tweets with a similar informative content and each group has been characterized with the most representative words appearing in tweets

III. Twitter Data Collection and Processing

A. K-mean Clustering

The K-mean method describes the following: if a dataset is *D* which encompasses n objects of observation, partition of distributing the object of observation in *D* into *K* clusters which are *C1,C2,C3,…,Ck.,* whereby $C_i \subset D$ and $C_i \cap C_j = \emptyset$. $(1 \leq i, j \leq k)$ [5]. Conceptually, the partition techniques require the centroid of a cluster which is known as the centre of points. Let x represents a cluster. The centre points can be defined as the mean of the object or points assigned to the cluster. The difference between Indonesian bullying words is measured by the distance between two points between x and y whereby is known as is Euclidian distance. The formula to calculate between all objects in x was introduced by Han, et al. [5].

$$E = \sum_{x=1}^{k} = \sum_{x \in y} dist(x, y)2 \quad (1)$$

E is a sum of the squared error for all Indonesian bullying words in the data set; y is the points in space ; and x is the centroid of a cluster (both x and y are multidimensional). The distance from the object to its cluster centre is squared and the distances are summed.

The algorithm k-mean is as follows:

- The dataset is divided into K clusters and the data point of Indonesian bullying words is allocated randomly in clusters until the clusters have the same number of data points.
- Each data point is calculated based on the distance from the data point to each cluster. If the calculation of the data points is too close to the own cluster, so the data points should not be removed, but when the data point is far from the own cluster, then the data point can be moved to near other clusters. If the results from a calculation of data points are similar to one another, it will be grouped into one cluster and if the results from a calculation of data points are far than the others, then it will be reallocated to the other clusters to be grouped.
- The step process above must be repeated until the data points do not move from one cluster to others. At these points, the clusters are stable and the clustering process ends.

B. Data Collection

The process of collecting the data is the same with the previous works; however the technique of data process is different from the previous works [3].

Figure 1 shows the process of data collection from Twitter and cleaning data process by using Rapid Miner. Extracting the textual message is the first process by retrieving Indonesian bullying key words.

The next process is cleaning data from tweets which are unstructured words. Tweets are transformed using Microsoft Excel format. This work only focuses on Indonesian bullying words. Tweets that do not contain Indonesian bullying words are removed.

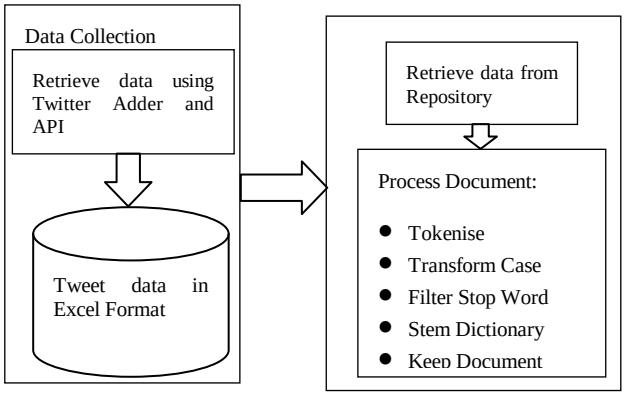


Fig. 1 Data cleaning and Processing Steps

The main characteristic of data sets that have been downloaded from Twitter is about 14997 tweets which specifically contain Indonesian bullying words, and the whole tweet messages that had been analysed. The data are gathered by establishing and maintaining a continuous connection. We monitored the public stream end points offered by Twitter Adder and Twitter APIs over a week and tracked a selection of keywords ranging over two different topics, i.e., rascal and stupid.

The data that have been captured contain tweets' ID, location, follower, friends, last tweets and last tweet date. This work analysed the last tweet and followers because the last tweet consists of Indonesian bullying

C.Data Processing using Rapid Miner

Within the analysis in relation to the data of Indonesian bullying words, researchers have to consider two attributes, i.e.; followers and tweets. The purpose of this is to know the extent the users' followers are influenced to write bullying message.

The calculation occurs between the scale value of follower and tweet attributes. The number of followers has to be divided by its number of followers set of 10,000 due to the fact that the amount of followers for each user varies. For instance, the largest number of followers is around 10,000, so all the value of users' followers is divided by 10,000. The objective from this division is to enhance the variable weights in favour of obtaining the best clustering through minimizing the ratio of the amount of cluster distortion over the amount of between-cluster distortion.

We also expected that the objective values in the followers attribute will a have smaller value which calculates the distance between tweets and followers attributes. The sum of followers and tweets can be determined using the formula of K-mean clustering.

Meta Data View | Data View | Plot View | Advanced Charts | Annotations

ExampleSet (14976 examples, 1 special attribute, 30 regular attributes)

Row No.	text	followers	friends			anjing	babi
1	sarap anjing goblok jancuk monyet setan babi bangsat	302	222	1	1	1	
2	gila gila anjing bangsat	148	179	1	1	0	
3	sarap anjing goblok jancuk monyet setan babi bangsat anjing	447	1133	1	1	1	
4	anjing goblok jancuk monyet setan babi bangsat anjing anj	776	103	1	1	1	
5	bangsat tolol anjing setan	485	973	1	1	0	
6	anjing bangsat monyet	351	419	1	1	0	
7	bangsat anjing	149	175	1	1	0	
8	anjing bangsat	162	801	1	1	0	
9	anjing bangsat	245	207	1	1	0	
10	bangsat anjing bangsat jancuk	190	84	1	1	0	

Fig. 2 Result of cleaning data from processing document

Figure 2 illustrates the result of data cleaning. The data displays Indonesian bullying words and followers that occurred in datasets.

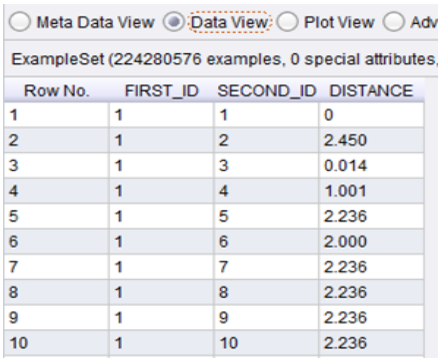
The data matrix in Figure 2 shows the frequency of Indonesian bullying words, followers and tweets. The row number and ID reflects the number of the tweet, for example, "bangsat" (rascal) term occurs in row number 1,2,3,…, etc.. The 0 means that the word does not appear in a tweet and 1,2,3 – a non-zero number means how many times the word appears in the tweet.

IV. Data Gathering and the Analysis of Finding

Mining Indonesian bullying words is the main purpose of this research. To achieve this, a clustering method is applied to analyse Indonesian bullying words which occur in Indonesian Twitter posts.

This work uses the K-mean technique as a method to analyse Indonesian bullying words including its followers and tweets in a set of points. The K-mean technique analyses Indonesian bullying words, followers and tweets by partitioning the words into a set of points. The purpose of this partition is to find some groups of Indonesian bullying words that involve their followers which have similar and dissimilar values. To find the similar and dissimilar values, K-mean clustering calculates the distance between Indonesian bullying words in tweet messages in datasets. The distance between Indonesian bullying words in tweets and followers can be defined using the Euclidean distance.

Before generating K-mean cluster, our work uses Rapid Miner to calculate the distance between Indonesian bullying words and the results are displayed in Figure 3.


Meta Data View | Data View | Plot View | Adv

ExampleSet (224280576 examples, 0 special attributes,

Row No.	FIRST_ID	SECOND_ID	DISTANCE
1	1	1	0
2	1	2	2.450
3	1	3	0.014
4	1	4	1.001
5	1	5	2.236
6	1	6	2.000
7	1	7	2.236
8	1	8	2.236
9	1	9	2.236
10	1	10	2.236

Fig. 3 Calculating the distance between Indonesian bullying words in Rapid Miner

Figure 3 provides the calculation of the distance between Indonesian bullying words which occur in tweets. The example of a calculation multi-dimensional between the first and second messages which contain Indonesian bullying words by using Euclidean distance is as follows:

$$d(x, Y) = \sqrt{(y1-X1)^2 + (y2-X2)^2 + ... + (yn-Xn)^2} \quad (2)$$

x and y represent the first and the second messages in the database, then the calculation between the first and the second messages is as follows:

$$d(x, y) = \sqrt{(y1-X1)^2 + (y2-X2)^2 + ... + (yn-Xn)^2}$$

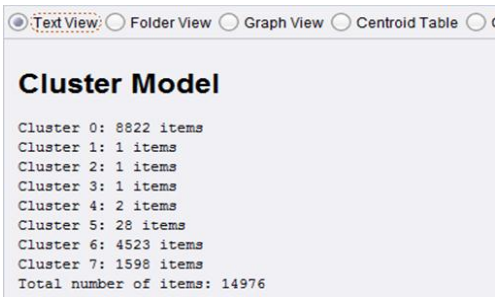
$$d(x,y) = \sqrt{(0.0148-0.0302)^2 + (1-0)^2 + (1-1)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2 + (1-1)^2}$$

$$d(x,y) = \sqrt{6.00023716}$$

$$d(x,y) = 2.45 \quad (3)$$

The distance between the first and the second messages is 2.445.

After calculating the distance between messages, we generate K-mean cluster using Rapid Miner. The result after generating K-mean clustering is shown in Figure 4: we divide K cluster into 8 clusters because we remove some clusters which do not have objects. Furthermore, we also remove some clusters which have value 0.

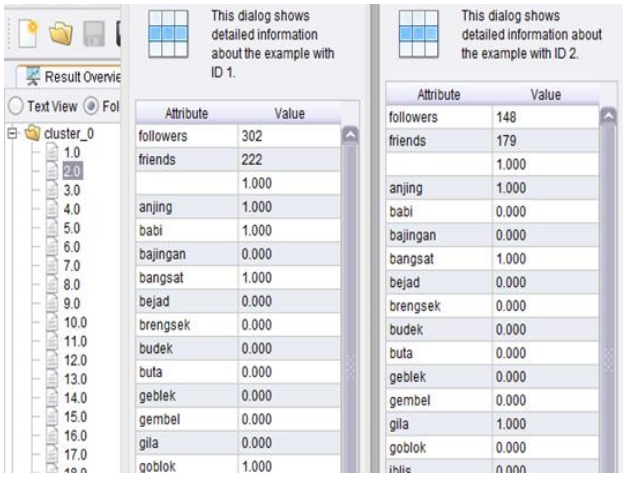

Text View | Folder View | Graph View | Centroid Table

Cluster Model

```
Cluster 0: 8822 items
Cluster 1: 1 items
Cluster 2: 1 items
Cluster 3: 1 items
Cluster 4: 2 items
Cluster 5: 28 items
Cluster 6: 4523 items
Cluster 7: 1598 items
Total number of items: 14976
```

Fig. 4 Result after generating cluster algorithm in Rapid Miner

Figure 4 shows that Cluster 0 has the largest amount of similar objects. Cluster 0 has 8822 items which means there are 8822 similar Indonesian bullying words in some messages. The followers are spread over 14976 tweets. Cluster 6 is the second large cluster which contains 4523 items.



Attribute	Value
followers	302
friends	222
	1.000
anjing	1.000
babi	1.000
bajingan	0.000
bangsat	1.000
bejad	0.000
brengsek	0.000
budek	0.000
buta	0.000
geblek	0.000
gembel	0.000
gila	0.000
goblok	1.000

Attribute	Value
followers	148
friends	179
	1.000
anjing	1.000
babi	0.000
bajingan	0.000
bangsat	1.000
bejad	0.000
brengsek	0.000
budek	0.000
buta	0.000
geblek	0.000
gembel	0.000
gila	1.000
goblok	0.000

Fig. 5 the example of cluster after generating K-mean cluster in Rapid Miner

Figure 5 shows one of the examples in cluster 0 where a user which has no more than 1000 followers, uses similar words in sending their offensive messages. The K-mean clustering method classified the Indonesian bullying words into categories which have similar definitions and a similar amount of followers.

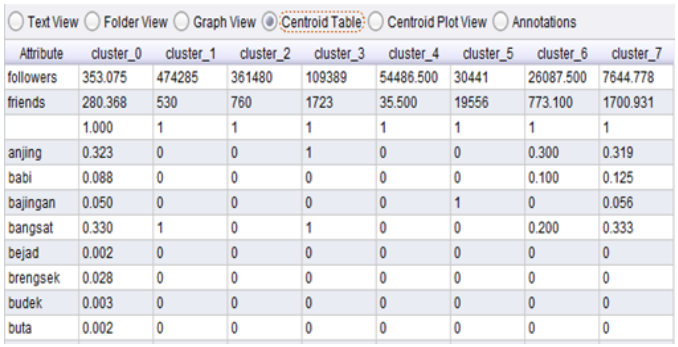
According to the results, after generating the k-mean algorithm, each cluster has their own characteristics which are shown below:

1) Cluster 0 has the largest amount of offensive messages compare to the other 7 clusters. In the dataset, cluster 0 has also a similar amount of followers. The feature from cluster 0 is a group of Indonesian bullying words associated with physical disability (disable person) and the area of psychology. Furthermore, the feature of this cluster has also a quantity of followers less than 1000.
2) Cluster 1. Based on the result of generating K-mean algorithm in Rapid Miner, cluster 1 contains a user's ID which have 474,285 followers, and an insulting message that is bangsat.
3) Cluster 2. The typical characteristic of cluster 2 is a user's ID that has 361,480 followers which is a little closer to cluster 1.
4) Cluster 3. Cluster 3 has a large number of followers (more than 100,000 followers). But, the combinations of insulting words in the message have no more than two insulting words.
5) Cluster 4. Cluster 4 has two users' IDs, typically 50,000 followers. Another typical feature of this is cluster 4 is that it has more than two insulting words.
6) Cluster 5. This cluster has only one user which has 30,441 followers. The bullying message in this cluster has one or two Indonesian bullying words.

7) Cluster 6. Cluster 6 hardly has 10,000 to 20,000 followers. There are only two insulting words being used in the message.
8) Cluster 7. Cluster 7 is the second largest group that has between 1000 and 10,000 followers.

From all these typical characteristics in each cluster, the conclusion that can be made is that even if the user has a large number of followers, not all of their followers will follow the user to send offensive and insulting messages to the victim. Users that have a large number of followers do not always influence their followers to send bullying messages. On the other hand, users that have followers below 1000 typically have similarities in sending bullying messages. This means that messages are being sent from one follower to another.

In addition to the result of calculation using K-mean clustering, it can also be seen in a centroid table. Figure 6 illustrates the distribution of the insulting words which contain Indonesian bullying words in a centroid table.



Text View | Folder View | Graph View | Centroid Table | Centroid Plot View | Annotations

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5	cluster_6	cluster_7
followers	353.075	474285	361480	109389	54486.500	30441	26087.500	7644.778
friends	280.368	530	760	1723	35.500	19556	773.100	1700.931
	1.000	1	1	1	1	1	1	1
anjing	0.323	0	0	1	0	0	0.300	0.319
babi	0.088	0	0	0	0	0	0.100	0.125
bajingan	0.050	0	0	0	0	1	0	0.056
bangsat	0.330	1	0	1	0	0	0.200	0.333
bejad	0.002	0	0	0	0	0	0	0
brengsek	0.028	0	0	0	0	0	0	0
budek	0.003	0	0	0	0	0	0	0
buta	0.002	0	0	0	0	0	0	0

Fig. 6 the centroid table after generating k-mean clustering using Rapid Miner software

Figure 6 describes that there are some clusters which enclose some Indonesian bullying words. The word “anjing” (dog) is spread across clusters 0,3,6 and 7 where each one of them is calculated by the K-mean algorithm. “anjing” (dog) in cluster 0 has the value of 0.33, one in cluster 3, 0.300 for cluster 6 and 0.319 in cluster 7. This shows that the word “anjing” (dog) is mostly located in cluster 3 because it has a value of 1, as well as another Indonesian bullying word. In the same way, “bangsat” (rascal) is in cluster 1 and 3 too, for the value is one. Whereas “keparat” which has 1 for its value is located in cluster 4, cluster 8 and cluster 5. Based on Figure 6, the categorization of Indonesian bullying words is positioned in cluster 0, cluster 6 and cluster 7. Cluster 0 is the particular group that has the largest number of Indonesian bullying words.

V. Conclusion

This paper introduces the analysis of the Indonesian bullying words using K-mean cluster. The purpose of clustering Indonesian bullying message is to find similarities and dissimilarities in the Indonesian bullying words in a group. The result of this work contributes important information about Indonesian bullying words to society, Indonesian government and non-government organizations.

Our preliminary experimental evaluation, performed on K-mean clustering shows that the effectiveness of the approach in discovering cluster between Indonesian bullying words is an interesting knowledge. The result of this work is to discover some groups of Indonesian bullying words which have similarities in words. Furthermore, most of messages which have a linkage to each other tend to make a group in a cluster. One cluster usually contains some messages coming from some people who have a relationship with the perpetrator or the victim. Most people who are involved in a group are usually follower perpetrator in Twitter.

Other interesting future research directions for further improvement regarding the performance of our work will be further developed. Furthermore, classifying Indonesian bullying words using the Naive Bayes method will be applied in future work.

Reference

[1] Pandaya, "Bullying is rampant in local schools: Survey," in *The Jakarta Post*, ed. Jakarta: The Bina Media Tenggara 2012.
[2] Tomothy, "Akamai: DDoS attacks increased since Q1 2013, Indonesia marked as biggest cyberbully," ed: engadget, 2013.
[3] H. Margono, X. Yi, and G. Raikundalia, "Mining Indonesian Cyber Bullying Patterns in Social Networks," in *ACSC*, 2014, pp. 115-124.
[4] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, pp. 651-666, 2010.
[5] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*, 3rd ed. Amsterdam: Elsevier, 2012.
[6] T. Xu and D. W. Oard, "Wikipedia based topic clustering for microblogs," *Proceedings of the American Society for Information Science and Technology*, vol. 48, pp. 1-10, 2011.
[7] K. D. Rosa, R. Shah, B. Lin, A. Gershman, and R. Frederking, "Topical clustering of tweets," *Proceedings of the ACM SIGIR: SWSM*, 2011.
[8] J. Weng and B.-S. Lee, "Event Detection in Twitter," in *ICWSM*, 2011.
[9] E. Y. Bann, "Discovering Basic Emotion Sets via Semantic Clustering on a Twitter Corpus," *arXiv preprint arXiv:1212.6527*, 2012.
[10] K. Chen and H. Zhao, "Twititude: Message Clustering and Opinion Mining on Twitter," 2012.
[11] E. Baralis, T. Cerquitelli, S. Chiusano, L. Grimaudo, and X. Xiao, "Analysis of Twitter Data Using a Multiple-level Clustering Strategy," in *Model and Data Engineering*, ed: Springer, 2013, pp. 13-24.