Third International Conference on “Materials, Computing and Communication Technologies”
in association with International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

# Twitter Post Approach for Feature Extraction and Communicating Metadata

**K. Amutha MCA. M.Phil., R. Janaki Rama**
[1]Assistant Professor, Department of Computer Science, Sri Sarada College for Women (Autonomous), Tirunelveli, Tamil Nadu, India
[2]II M.Sc. Computer Science, PG and Research Department of Computer Science, Sri Sarada College for Women (Autonomous), Tirunelveli, Tamil Nadu, India

## ABSTRACT

Because of the enormous use of social networking, Twitter has become a popular medium for disseminating information. According to existing studies, Twitter can be utilized for successful post-incident measures since it has a considerably greater reach than conventional media. On Twitter, people employ colloquial language such as acronyms, misspelling words, synonyms, transliteration, and unclear terminology. As a result, extracting incident-related data is a difficult undertaking. This information, on the other hand, may be useful to public safety groups in the event of an emergency. This study presents a framework for extracting and reporting early event-related information from Twitter streams, which monitors and synthesizes event-specific information, such as a terrorist attack, and warns law enforcement, emergency services, and media sources. Tweet-to-Act (T2A) is a suggested system that uses word embedding to convert tweets into a vector space model and then uses the Word Mover's Distance (WMD) to cluster tweets for event detection. The proposed system uses sequence labelling with random forest to extract trustworthy and relevant information from a huge dataset of short and informal tweets

**INDEX TERMS** : Random forest, machine learning, ,Sentiment analysis, Natural Language processing, opinion mining

## I. INTRODUCTION

Microblogging has been a major study field for sentiment analysis. In platforms like Twitter, people submit messages about their everyday lives, and these tweets are often linked to a wide range of themes. Prior to and throughout the feature selection process, pre¬processing is critical to the success of many research on sentiment analysis classification algorithms. Texts are cleaned and prepared for classification as part of the pre-processing process.

On the Internet and in particular on Twitter, there is a large quantity of noise. Data that does not provide any helpful information for our study, such as sentiment analysis, is referred to as "noise."

As much as 40% of a dataset has noise, which presents problems for machine learning systems. The use of acronyms and slang is common among Twitter users, as are spelling and typographical mistakes. [1] – [3]They may also employ punctuation marks, such as exclamation marks, to express their feelings. Most of the time, it is not required to incorporate every word in the original form of the text in the machine learning stage. To ensure that machine learning succeeds, pre-processing data must be cleansed and normalized, which necessitates this step.

Using two well-known datasets, this work aims to collect and assess a variety of popular pre-processing methods as well as some innovative ones, such as the replacement of contractions and negations with antonyms, to see whether they have any importance in feature selection. Based on our findings, we recommend to future researchers which strategies are most suited for Twitter sentiment analysis, and which ones should be avoided.

Text sentiment analysis begins with pre-processing, which may be made more accurate by using several methodologies. We tried a large number of pre-processing methods that had never been compared before on two different datasets. Three example machine learning algorithms assessed the accuracy of each strategy. Finally, depending on the findings, we divide the techniques into different performance groups and count the number of features for each one. Tests demonstrate that certain Twitter sentiment analysis algorithms are more accurate than others when it comes to classifying tweets, while others are less reliable. Stemming, replacing punctuation repeats, and deleting numerals are all highly suggested.

## II. EXISTING SYSTEM

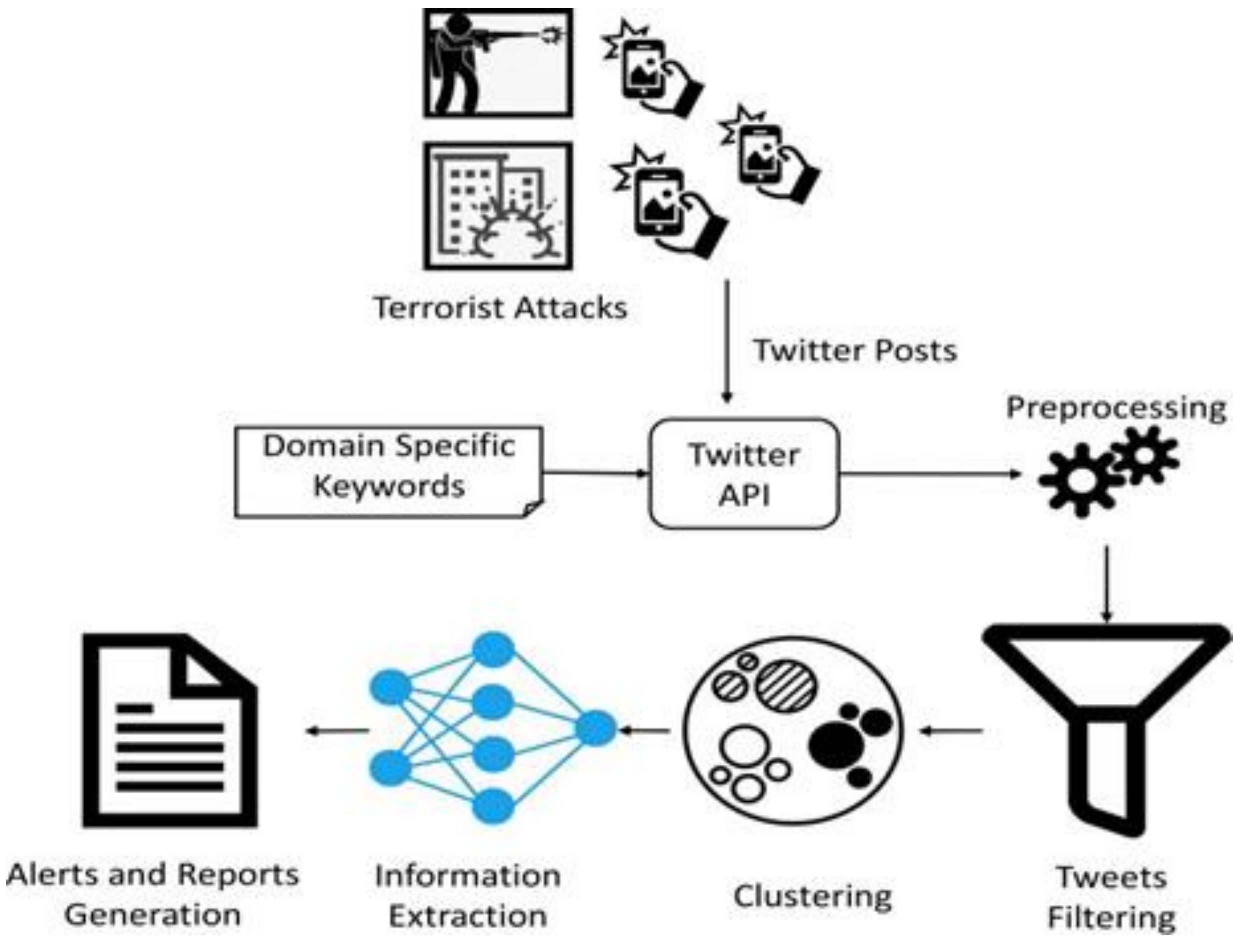
Online Social Networks (OSNs) such as Facebook, Google, and Twitter are inherently designed to enable people to share personal and public information and make social connections with friends, co-workers, colleagues, family, and even with strangers. [3]-[6] In recent years, we have seen unprecedented growth in the application of OSNs. For example, Twitter, one of representative social network sites, claims that it has more than 800 million active users and over 30 billion pieces of content (web links, news stories, Tweets, blog posts, notes, photo albums, and so on.) shared each month. To protect user data, access control has become a central feature of OSNs. A typical OSN provides each user with a virtual space containing profile information, a list of the user's friends, and webpages, such as wall in Twitter, where users and friends can Tweet content and leave messages.

A user profile usually includes information with respect to the user's birthday, gender, interests, education, work history, and contact information. In addition, users can not only upload a content into their own or others' spaces but also tag other users who appear in the content. Each tag is an explicit reference that links to a user's space. For the protection of user data, current Twitter indirectly require users to be system and policy administrators for regulating their data, where users can restrict data sharing to a specific set of trusted users. OSNs often use user relationship and group membership to distinguish between trusted and untrusted users. For example, in Twitter, users can follow peoples, create friends, groups, or public to access their data, improve our business qualities and income, depending on their personal authorization and privacy requirements.

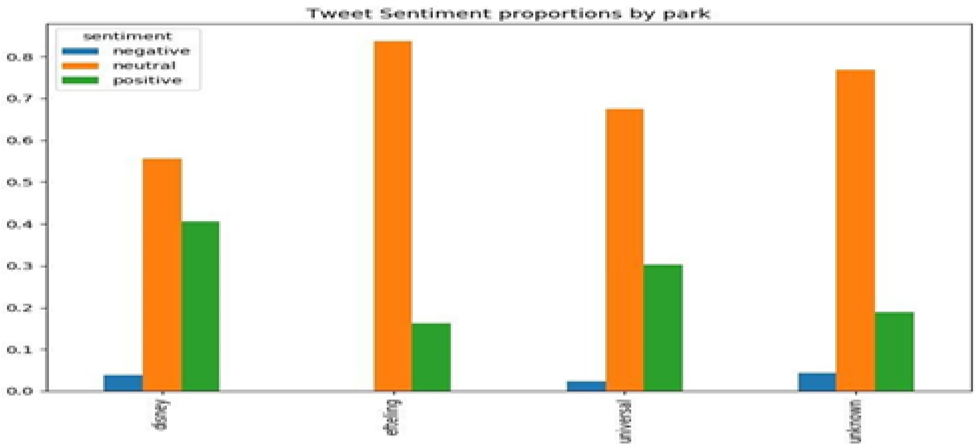
## III. PROPOSED SYSTEM

In Proposed System this project is implemented with a proof-of-concept Twitter application for the collaborative management of shared data, are managed by sentiment analysis. Our prototype application enables multiple associated users to specify their authorization policies and privacy preferences to CNN based co- control a shared data item. It is worth noting that our current implementation was restricted to handle photo sharing in OSNs. our approach can be generalized to deal with other kinds of data sharing and comments, in OSNs as long as the tweets holder of shared data are identified with effective methods like tagging or searching.

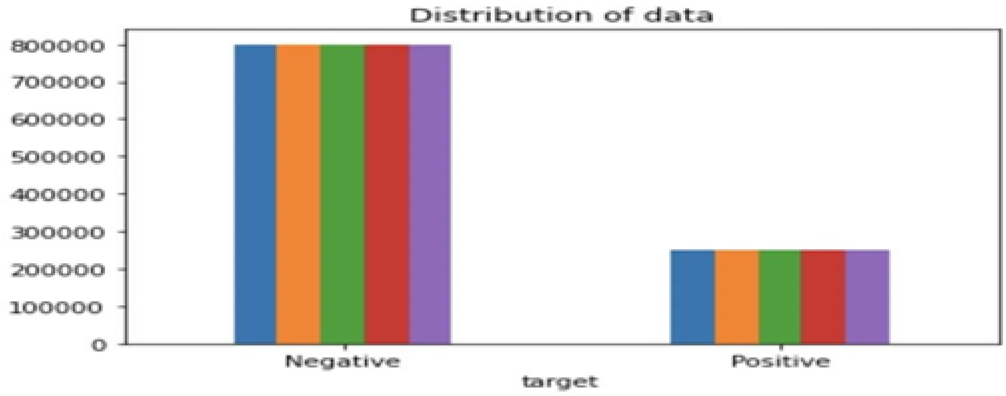
The proposed system shows a novel solution for collaborative management of shared data in OSNs. We have introduced an approach to propose and experimentally evaluate an automated system, called Filtered Wall, able to filter unwanted messages from OSN user walls. This project exploit Machine Learning text categorization techniques to automatically assign with each short text messages a set of categories based on its content.[7]-[9] A system to automatically filter unwanted tweet messages or emojis from OSN user walls on the basis of both message content and the message creator relationships and characteristics.



## IV. RESULT




In this figure it shows that tweet sentimenrt proportions by park instead of many data's are written in pseudo assembler in negative ,positive and neutral approaches




In this figure it shows distribution of data's of many data's are read in negative and positive approaches are milliseconds counts

## V. CONCLUSION

As Compared to the support vector machine and the baseline method for Twitter sentiment classification algorithm, the results indicate that the random forest has obvious advantages in datasets. This shows that random forest can effectively construct text semantics. The random forest directly models context sentiment feature from text, and selects the most important features in the tweet effectively. It avoids error propagation and improve the classification performance. The pre-trained word vector representation by learning on Twitter corpus can better describe the similarity between words in Twitter, and extract the implicit semantic relation and sentiment feature information between words in the tweet.

The similarity between words in Twitter, and extract the implicit semantic relation and sentiment feature information between words in the tweet. This project presented a comprehensive set of experiments for both these tasks on manually annotated data that is a random sample of stream of tweets. We investigated two kinds of models: tree kernel and feature based models and demonstrate that both these models outperform the unigram baseline. For our feature-based approach, this project proposes feature analysis which reveals that the most important features are those that combine the prior polarity of words and their parts-of-speech tags.

## VI.REFERENCES


[1]. Socher R et al., "Recursive deep models for semantic compositionality over a sentiment treebank," in Proc. Conf. Empirical Methods Natural Lang. Process., Seattle, DC, USA, Oct. 2013, pp. 1631-1642

[2]. Fern M, He Y and Saif H, "Evaluation datasets for Twitter sentiment analysis: A survey and a new dataset, the STS-Gold," in Proc. 1st ESSEM Workshop, Turin, Italy, 2013, pp. 21-26.

[3]. FuruWei Y N, Liu T, Qin B, Tang D and Zhou M, "Learning sentiment specific word embedding for twitter sentiment classification," in Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, Baltimore, MD, USA, Jun. 2014, pp. 1555-1565,2014.

[4]. FuruWei Q B, Liu T, Tang D and Zhou CoooolllM:, "A deep learning system for twitter sentiment classification," in Proc. 8th Int. Workshop Semantic Eval. (SemEval), Dublin, Ireland, Aug. 2014, pp. 208-212.

[5]. Martíín-Valdivia M T, Martínez-Cámara E, Montejo-Rááez A and Ureña-López L A, "A knowledge-based approach for polarity classification in Twitter,'' J. Assoc. Inf.Sci. Technol., vol. 65, no. 2, pp. 414-425, 2014.

[6]. Jianqiang Z and Xueliang C, "Combining semantic and prior polarity for boosting twitter sentiment analysis,'' in Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom (SmartCity), Dec. 2015, pp. 832-837.

[7]. Johnson R and Zhang T, "Semi-supervised convolutional neural networks for text categorization via region embedding," in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 919-927.

[8]. Büüchner M, Hagen M, Potthast M and Stein B, "Twitter sentiment detection via ensemble classification using averaged confidence scores," in Proc. Eur. Conf. Inf. Retrieval, 2015, pp. 741-754.

[9]. Jianqiang Z and Xiaolin G, "Comparison research on text preprocessing methods on twitter sentiment analysis,'' IEEE Access, vol. 5, pp. 2870- 2879, 2017.

[10].M. A. Gilbert, "Strengthening your social media marketing with live streamingvideo,''inSmartTechnologiesandInnovationforaSustainable Future. Cham, Switzerland: Springer, 2019, pp. 357–365.