



Multivariate Forecasting using Natural Language Processing to Gain insights of Social Media Data

Dr.Ch. Surya Kiran¹ | Perni Naga Pavani² | Syed Afra² | Lankapalli Suma² | Mullangi Priya²

¹Professor, Department of CSE, NRI Institute of Technology, Vijayawada, A.P., India.

²Department of CSE, NRI Institute of Technology, Vijayawada, A.P., India.

To Cite this Article

Dr.Ch. Surya Kiran, Perni Naga Pavani, Syed Afra, Lankapalli Suma and Mullangi Priya. Multivariate Forecasting using Natural Language Processing to Gain insights of Social Media Data. International Journal for Modern Trends in Science and Technology 2022, 8(03), pp. 99-101. <https://doi.org/10.46501/IJMTST0803018>

Article Info

Received: 08 February 2022; Accepted: 10 March 2022; Published: 16 March 2022.

ABSTRACT

Detection and classification of events in social media is different from detection and classification in other text format. The rationale why detection of events in social media streams is tougher is brief and noisy content, diverse and rapidly changing topics, and enormous volumes of knowledge. Many people within the world use social media to remain connected to their friends, relations and colleagues through their computers and mobile phones. Due to their real-time existence, social media like Facebook, Snapchat, Whatsapp and Twitter have recently received tons of attention. The proliferation of social media exposes many opportunities for research. For several reasons, social media information might be used, like monitoring accidents, predicting events, and even early warning systems. Messages posted on Twitter revealed everything from everyday tales to the new news and events at the local and global level. Event detection and classification has become a crucial task in social media platforms. It facilitates the exploration and navigation of events with early preventive action plans. The most challenges are the features of short / conversational, heterogeneous and live social media data. Online social network apps like Facebook, Weibo, have played a key role within the lives of individuals. Now days there's rapid development of social media platforms and Twitter is one among the social media's most famous platforms. This project focuses on collecting numerous tweets from twitter, preprocessing the tweets and therefore the tweets are classified into specific categories by using Decision tree classifier.

1. INTRODUCTION

Detection and classification of events in social media is different from detection and classification in other text format. The reason why detection of events in social media streams is more challenging is short and noisy content, diverse and rapidly changing topics, and large volumes of data. Millions of people in the world use social media to stay connected to their friends, family members and colleagues through their computers and mobile phones. Because of their real-time existence, social media like Facebook, Snapchat, Whatsapp and Twitter have recently received a lot of attention. The proliferation of social media opens up many

opportunities for research. For many reasons, social media information could be used, such as monitoring accidents, predicting events, and even early warning systems. Messages posted on Twitter (tweets) revealed everything from everyday tales to the latest news and events at the local and global level.

Event detection and classification has become an important task in social media platforms. It facilitates the exploration and navigation of events with early preventive action plans. The main challenges are the features of short / conversational, heterogeneous and live social media data. Online social network apps like Facebook, Weibo, have played a key role in the lives of

people. Tweets contain tremendous data. But it's a difficult problem how to mine the tweets and get valuable information. This paper focuses on collecting numerous tweets from twitter, preprocessing the tweets and the tweets are classified into specific categories by using Decision tree classifier.

EXISTING SYSTEM

Some of the current popular classifiers are support vector machine (SVM), neural network (NN), KNN, Naïve Bayes and so on, are built in an inductive learning way. Among them, SVM is acclaimed by many researchers for its leading performance. Therefore, it has been widely used for text classification purpose. The existing model for statistical analysis classifies the data into specified category by using the classification algorithm Support Vector Machine (SVM).The execution time is more and efficiency and accuracy are less when compared with Decision tree classifier.The structure of algorithm is very complex.

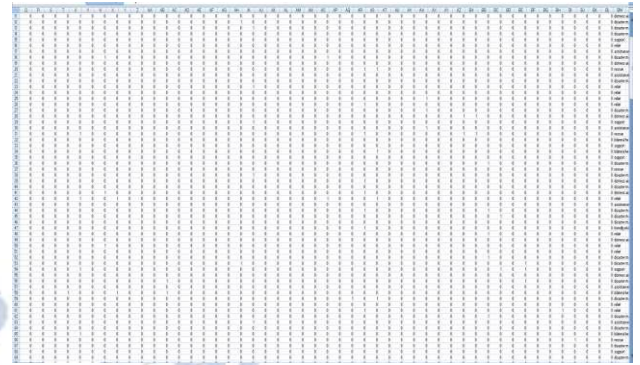
2. PROPOSEDSYSTEM

We propose a model for statistical analysis of data that uses "Decision Tree" (DT) a machine learning algorithm instead of SVM and Naïve Bayes. Decision Tree is nonlinear classifier and it is better in terms of accuracy . It is very easy to implement and understand. It has very good performance and it is fast.

3. IMPLEMENTATION

Here we consider some tweets from the twitter and by using an algorithm decision tree classifier we will classify the twitter tweets.First we consider a tweet based on the keywords that we use it classifies the tweet that means we give some keywords in dataset so based on that it compares and classifies the given tweet into their respective domain.

SAMPLE SCREENS



```

disaster management 20
distress aid 2
bilateral help 2
rescue 2
relief 2
support 1
friendly relations 1
dtype: int64
the accuracy with Naive Bayes Classifier is:
0.03333333333333333

```

```

the accuracy with Decision Tree Classifier is:
0.4
disaster management 17
relief 6
distress aid 5
assistance 2

```

4. CONCLUSION

We collect tweets from twitter and those tweets will be preprocessed.Natural Language Processing Decision tree classifier is used to classify the data into their respective domains.Decision tree classifiers have good accuracy.They are so useful to get aware of the things using tweets.

FUTURE SCOPE FOR FURTHER DEVELOPMENT

The conducted experiments showed a good performance of Decision Tree Classifier and achieved a good overall accuracy. In future accuracy of the same can be improved with the help of neural networks.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

- [1] Twitter Trending Topic Classification Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md. Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary Department of Electrical Engineering and Computer Science Northwestern University, Evanston, IL 60208 USA Email: {kml649, drp925, ran310, mpatwary, ankitag, choudhar}@eecs.northwestern.edu
- [2] W. Zhang and F. Gao, —Procedia Engineering An Improvement to Naive Bayes for Text Classification, || vol. 15, pp. 2160–2164, 2011.
- [3] Short Survey on Naive Bayes Algorithm Pouria Kaviani, Mrs. Sunita Dhotre M.Tech student, Department of Computer Engineering, Bharati Vidyapeeth University, College of Engineering, Pune Associate Professor, Department of Computer Engineering, Bharati Vidyapeeth University, College of Engineering, Pune.
- [4] P.Selvaperumal and A. Suruliandi, "A short message classification algorithm for tweet classification," 2014 International Conference on Recent Trends in Information Technology, Chennai, 2014, pp. 1-3.
- [5] Natural Language Processing: State of The Art, Current Trends and Challenges Desha Khurana, Aditya Koli, Kiran Khatter and Sukhdev Singh Department of Computer Science and Engineering Manav Rachna International University, Faridabad-121004, India Accendere Knowledge Management Services Pvt. Ltd., India