

## CATEGORIZAION OF DRUGS BASED ON POLARITY ANALYSIS OF TWITTER DATA

**Banasmita Borah, M.Phil Final Year,  
Department of Computer Science,  
Mother Teresa University,  
Chennai- 15**

**Dr. Mrs. R. Janaki, MCA., M.Phil., Ph.D.,  
Assistant Professor  
Department of Computer Science,  
Queen Mary's College (A),  
Chennai-600 004**

**Abstract**— In today's world, all the folks are unit terribly strung-out in social networking. The main objective of the system is to gather medical tweets from Twitter supported the patient's feedback. This technique can predict the most effective drug mistreatment polarity prediction. All the data from the social networking websites may be used effectively. This technique can analyse the tweets data and provides out the most effective drugs for the buyer within the existing system, it assist Food and Drug Administration (FDA) in distinguishing ADR (Adverse Drug Reaction) messages on net forum and result may be used as early warning system. By this technique, the user won't recognize that is that the smart drugs. Within the projected system the buyer can gain data concerning the most effective medicines.

**Key words:** FDA (Food and Drug Administration), ADR (Adverse Drug Reaction).

### I. INTRODUCTION

Big knowledge is high-volume, high-speed, high truthfulness, high worth and high-variety info assets that demand cost-efficient, innovative styles of informatics for increased insight and decision-making. Huge knowledge sometimes includes knowledge sets with sizes on the far side the flexibility of ordinarily used software package tools to capture, create, and manage and method the info at intervals a tolerable period of time.

“CATEGORIZAION OF DRUGS BASED ON POLARITY ANALYSIS OF TWITTER DATA” reserves the right to do the final formatting of your paper.

### EXISTING SYSTEM

In the existing system, the physicians and pharmacists post the messages related to new medication that have discharged at intervals the market recently in web forum. Even the consumers unit allowed to post their experience once exploitation the medication. The system integrates every text and processing techniques to automatically extract very important text choices from the posts initial, Then classify the posts into positive/negative examples supported variety of pre-identified ADR

connected posts Filtering mechanism is completed exploitation text classification. The LDA is also a generative probabilistic model that uses a little selection topics to elucidate a gaggle of documents and it effectively crop the dimension of the texts. this technique assist Food and Drug Administration (FDA) in characteristic ADR (Adverse Drug Reaction) messages on web forum and result's used as early warning system.

## PROPOSED SYSTEM

The planned system uses Twitter to induce the info and technique on it. User can move the system through coder. New user have to be compelled to turn out associate account by giving the username and watchword, the registered user can directly login and may enter into the system twitter search space. In search space user can offer the input, and user get the tweets from the twitter exploitation twitter4j supported internet locomotion. The extracted tweets unit of measurement the preprocessed by removing stop words, short kind and emoticons and hold on in info. To extract the tweets, initial the association has to be compelled to be established with twitter account exploitation the twitter API called twitter4j. Then turn out the twitter developer application in twitter developer computing machine. The info from the Twitter is extracted exploitation locomotion and Twitter API. The twitter API will crawl the tweets from twitter exploitation twitter4j. Twitter4j will connect with Twitter exploitation designed Application .Twitter application is intended supported the customer key, secret key, Access token and token secret key. Exploitation these keys and tokens the association is established with Twitter and Twitter4j will extract the tweets and show into the user I the table format. These extracted tweets unit of measurement then preprocessed by substitution the short kind words with full kind. Eg: “r” is replaced with “are” , “2” is replaced with “TO”. It in addition replace the emoticons with its individual which suggests. Eg: #-:-) means Smiling with a hat. It in addition exclude the stop words kind the extracted tweets. Eg: “the”, “of”. These preprocessed tweets unit of measurement then hold on among the knowledge. The preprocessed tweets unit of measurement a lot of classified exploitation SVM classification based totally upon the category. Throughout this method its classified supported drugs connected tweets

and diseases connected tweets. Polarity detection is finished by the keywords like wise, bad etc. This technique is extraordinarily useful for the users to comprehend information regarding the best medication.



This diagram shows the connection between completely different elements of system. This diagram is extremely necessary to know the general conception of system. Design diagram could be a diagram of a system, within which the principal components or functions are delineate by blocks connected by lines that show the connection of the blocks. they're heavily utilized in the engineering world in hardware style, electronic style, software system style, and method flowchart. Architecture on polarity analysis of twitter data is used to analyse the best medicine. Initially, the user should login into the search page by giving user name and password if the user is a registered user otherwise the user has to register first, then login to search page. When user fill the information during registration it gets stored in database so during the login it gets validates from the database. If the username and password is correct then it enters into twitter search page. In twitter search page, the user can give the input to extract and crawler the tweets from the twitter. To extract the tweets, first we have to connect the application with twitter using Twitter API. Using Twitter4j the application get connected to twitter application based on consumer key, secret key, Access token and

token secret key. The preprocessed data is then stored in database. In future, the preprocessed tweets are classified using SVM classification. The polarity of the words such as good or bad etc is identified from classified tweets to classify whether it is positive tweets or negative tweets. Polarity can be predicted based on the number of positive comments and negative comments.

#### **TWITTER EXTRACTION**

User can act as interface between the user and conjointly the system. New user ought to produce Associate in nursing account by giving the username and information. The registered user can directly login and would possibly enter into the system twitter search space. In search space user can provide the input, and user get the tweets from the twitter. To extract the tweets, first the association got to be established with twitter account exploitation the twitter API called twitter4j. Then turn out the twitter developer application in twitter developer electronic computer. From the developed application we have a tendency to tend to induce the customer key, secret key, Access token and token secret key. Exploitation these keys and tokens, it's designed and connected with twitter. Throughout this API it contains many parameters to extract and skim from the Twitter manufactory by exploitation question search and need to maintain the question search ends up in question Result. Exploitation get Tweets technique we have a tendency to square measure ready to get the tweets, from that we have a tendency to square measure ready to extract the tweet username.

#### **PREPROCESSING**

The extracted tweets square measure the preprocessed by removing stop words, short kind and emoticons. All unmeaningful words within the tweets like stop words square measure been removed. All short forms are replaced with full words in order

that it's apprehensible for all the users. Emoticons square measure referred to as smileys; there square measure varies styles of smileys. for every smileys there square measure some emotional feelings in it, that the user use to speak in a lot of easier manner however it's not necessary all the user can recognize the which means of all emoticons. So, all the emoticons is replaced with their individual which means.

#### **SVM CLASSIFICATION**

Support Vector Machines square measure supervised learning models with associated learning algorithms that analyze information and acknowledge patterns, used for classification and multivariate analysis. Support Vector Machines square measure supported the conception of call planes that outline call boundaries. a call plane is one that separates between a collection of objects having totally different category memberships. A schematic example: medication and diseases. Once the Preprocessing the tweets square measure classified into diseases and medicines connected tweets. The words square measure known supported the keywords to classify the tweets. This lexicon analysis technique is employed to search out the well-liked class from the big range of tweets.

#### **POLARITY PREDICTION**

The classified tweets are analyzed supported polarity of the words like smart, bad, not, un etc. supported the polarity the quantity of positive tweets and negative tweets are known. we tend to are victimization the SVM classifier for classification technique for locating the polarity of the tweets and comments like positive tweets, negative, mixed or neutral.

## RESULT ANALYSIS

In the system could connect with Twitter by victimization Twitter application details through Twitter API. Tweets square measure being extracted so quickly from the Twitter victimization Twitter4j. These tweets square measure preprocessed to urge eliminate stop words, replace the short kind with full kind and replace the emoticons with its corresponding which suggests for the simple understanding for the users. In future work, it's additional classified and analyses the best medication victimization polarity. The input for the system need to tend in Twitter Search house to extract the tweets from the Twitter victimization Twitter API and Twitter application. Input are medication or diseases that the system will extract the tweets supported the given input. The tweets are getting to be displayed inside the table format throughout a number of seconds. The table will contain the username and their denote tweets. The system is analyzed before preprocessing and once preprocessing. In some cases range the amount the quantity of words in tweets before preprocessing are getting to be over range of words once preprocessing. In some cases range the amount the quantity of words in tweets before preprocessing are getting to be same as a result of the amount of words once preprocessing. In some cases quantity the quantity the number of words before preprocessing may be a smaller amount than the quantity of words once preprocessing. of those cases is based on

## REFERENCES

- [1] Bruno HS, Bruce MP. Detection, verification, and quantification of adverse drug reactions. *BMJ* 2004;329:44–7.
- [2] van Grootheest K, de Graaf L, de Jong-van den Berg LTW. Consumer adverse drug reaction reporting – a new step in pharmacovigilance? *Drug Saf* 2003;26:211–7.
- [3] van Hunsel F, Talsma

removal of stop words like “of”, “the” etc , replace for temporary kind with full words. Eg “u” is replaced with “you” and replace the emoticons with its corresponding which suggests. Eg “%-) “ is replaced with its which suggests “Confused or merry”.

## Conclusion

The planned system for categorizing medication supported polarity analysis of twitter information. From the twitter developed application all the keys and token square measure generated, with these data we will connect the twitter with twitter API. The twitter tweets square measure extracted with twitter API victimization twitter4j. Twitter4j can connect with Twitter victimization designed Application .Twitter application is designed supported the buyer key, secret key, Access token and token secret key. Then extracted tweets square measure preprocessed by removing stop words, short forms and emoticons. The preprocessed tweets square measure keep in info. The preprocessed tweets square measure any classified victimization SVM classification primarily based upon the class. during this system it's classified supported medication connected tweets and diseases connected tweets. Polarity detection is completed by the keywords like smart, bad etc. supported variety the amount the quantity of positive tweets and therefore the number of negative tweets it analyses the simplest medication.

A, van Puijenbroek E, de Jong-van den Berg L, vanGrootheest K. The proportion of patient reports of suspected ADRs to signal detection in the Netherlands: case-control study. *Pharmacoepidemiol Drug Saf* 2011;20:286–91.

[4] de Langen J, van Hunsel F, Passier A, de Jong-van den Berg L, van Grootheest K. Adverse drug reaction reporting by patients in the Netherlands – three years of experience. *Drug Saf* 2008;31:515–24.

- [5] Fernandopulle RBM, Weerasuriya K. What can consumer adverse drug reaction reporting add to existing health professional-based systems? Focus on the developing world. *Drug Saf* 2003;26:219–25.
- [6] Yang M, Li Y, Kiang M. Uncovering social media data public health surveillance. In: Proceedings of the 15th Pacific-Asia conference on information system, Brisbane; 2011. Paper 218.
- [7] Fichman RG, Kohli R, Krishnan R. The role of information systems in healthcare: current research and future trends. *Inform Syst Res* 2011;22:419–28.