International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 4109-4112


# Enhancement of Healthcare Using Naïve Bayes Algorithm and Intelligent Datamining of Social Media

**Suraj Kumar*, Aman Jain* and P. Mahalakshmi**
*Department of Computer Science, SRM Institute of Science and Technology,*
*SRM University, Chennai-603203 District- Kancheepuram, India.*
**Corresponding Author*

**Abstract:**
In this paper, we propose a novel quality mindful data digging programming empowered framework for cleverly gaining the knowledge about the cancer medicine, to educate and interact with patients. The Social media can also used for data extraction for simultaneously improving health care outcomes using user generated sentimental analysis of medicine to improve the quality of drugs by using social media as a resource. It is a great technology advancement in medical science. The heavy assurance on social networking sites led them to generate massive data on social media and which increase the opportunity of improvement of the medicine. Physician and biomedical scientist could collect the feedback from other doctors and patients to improve their knowledge enlightened for healthcare decisions. Naive Bayes algorithms are used for performing the sentimental analysis for differentiating the positive and negative reviews of the patients. This framework helps the medical institute for obtaining better knowledge for choosing which drug getting more benefit and give minimum side effects.

**Keywords:** sentimental analysis, Naive Bayes, Social media, healthcare decisions.

## INTRODUCTION

Social media is giving boundless opportunities for patients to confabulate their experience about the drugs aftermath and for companies to receive feedback on their products. Data mining is the way toward deciding the examples in colossal informational collections including techniques at the convergence of machine learning and coordination. In this framework data mining uses citified mathematical algorithm to reprocess data sets and gauge the probability of the positive and negative sentiments.

Opinion mining utilizes lingual handling, content investigation and computational phonetics to examine the full of feeling states and subjective data. Opinion mining is the process of resolving whether a piece of writing is positive, negative or neutral. It is also termed as sentimental analysis. Pharmaceutical organizations are organizing interpersonal organization observing inside their areas of expertise, building up an open door for fast scattering and criticism of items and administrations to upgrade and improve conveyance, development turnover and profit, and decline expenses. Data pre-processing is an imperative stamp in the data mining process. Data pre-processing involves the deportation of missing words, impossible data similarities, spoiled word etc. Data pre-processing involves data gathering, representation and quality of the data before analysis thus eliminating misleading of the data. Graphical portrayal is the most extreme ordinary strategy to outward speak to the measurements. The idea of informal communities makes information gathering troublesome. A few strategies have been utilized for example, interface mining, characterization through connections. Bar graph represents the sentimental analysis of various patients on a specific drug. The bar graph helps to understand the analysis of the drugs to the companies.

### Naïve Bayes:

Naive Bayes classification involves classifying of different models according to their special features. In this project Naive Bayes classifier classify the negative and positive sentiments of the patients. Naive Bayes had been used in "Machine Learning Algorithms for Opinion Mining and Sentiment Classification"[2]. In the former journal it has been used for textual classification of the text mining data, and they described the way to effectively use this technique for Opinion mining.

### Correlation identification between user post and their opinion:

Data correlation between the user and his post can be checked with coded data set in which data set have been cross checked with the companies verified data set which relates the data set to compare with SOM(Self Organizing Maps) graphs[1]. However, they have applied the method only to lung cancer while we plan to expand the method to various other cancer and anxiety related diseases. The utilization of National Library of Medicine's Medical Subject

Heading (MeSH)[1] which is a controlled vocabulary that comprises of a pecking order of descriptors and qualifiers that are utilized to comment on medicinal terms was done for tagging various medical terms into categorizing sentiments.

### Sentimental Analysis:

In this system Sentiment examination includes characterization of the informational collection as indicated by specific likenesses. Slant investigation or Opinion Mining is the computational treatment of assessments, estimations and subjectivity of content[3]. It is a Natural Language Processing and Information Extraction assignment that means to acquire essayist's emotions communicated in positive or negative remarks, inquiries and solicitations, by dissecting a huge quantities of archives[2],[3]. They have been controlled by utilizing Naive Bayes classifier. In the previous diary "Machine Learning Algorithms for Opinion Mining and

Sentiment Classification"[2] it has been arranged utilizing three different techniques.

**MODULES**

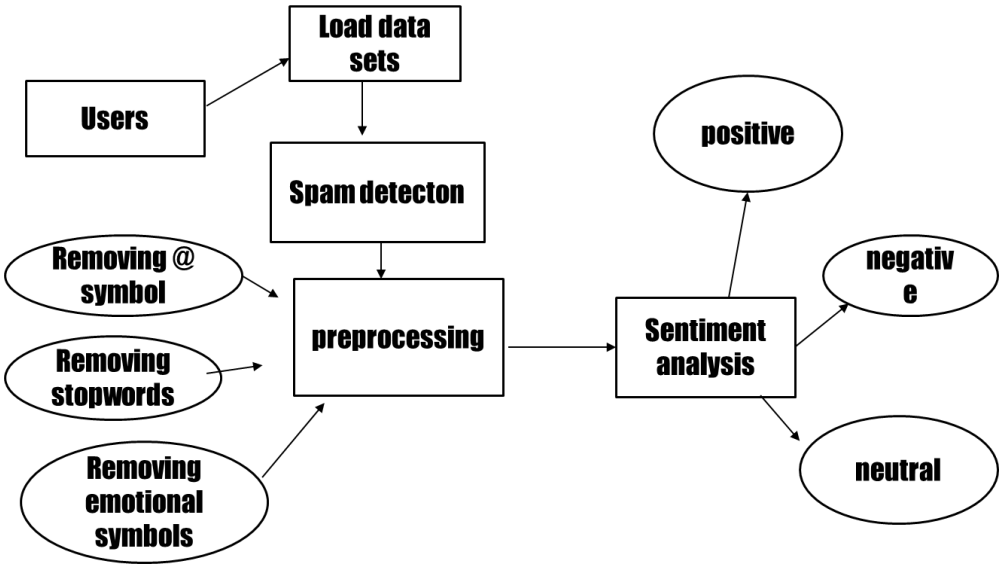


**Figure:** ER diagram shows the working of the project module

ER diagram demonstrate the working model of the framework and interconnection of the modules. The process start from the uploading of the data set into software and it is then followed by the Spam detection module which eliminates all the spam reviews. Again this process is forwarded to the pre-processing module where data such as @, stop words, emoticons are filtered from the data set. After the former module it went to the sentimental analysis module where we get the positive, neutral and negative reviews of the specified drugs based of the user's opinion. Graph is plotted according to the sentiment variation tracking that shows the effectiveness of the medicine.

**Initial Data Collection and Input Module:**

We input the data collected from websites in the form of Excel sheet as data sets into our software by loading it using My SQL Connection. The user will load the dataset in the software which comprises of Patient Name, Drug Name, and reviews on the drug by patients.

**Spam Detection Module:**

Once the dataset in input to the software, next we move on to Spam Detection module. This module is used for detecting incorrect and vague reviews amongst all other reviews and separating them apart. We have included medical dictionary in the software which will compare the side effects mentioned by users in the data set to the description of medicine and uses preprocessing methods to check for any discrepancies in the data. Such spam comments by user are deleted and removed from the data to preserve the integrity and correctness of software and remove any irrelevant comments.

**Pre-Processing Module:**

The next module that comes after Spam Detection is Pre-Processing. This part involves Removing of @ symbol, Stop words such as " is, a , it, less, ie, however, mostly, several, many, very, down, above, along, etc" etc. Also, preprocessing involves removing Emoticons to bring out weighted data from the input feedback(Ontology) based data. All this is done by comparing the data from users to certain classifiers defined by us in the software. Once weighted data is drawn out of the Ontology data, it can be effectively used for Sentiment Analysis for showing Users satisfaction/dissatisfaction in using certain medicines.

**Sentiment Analysis And tracking Module:**

Sentiment analysis is the estimation of positive and negative dialect. It is a critical process because it helps you see what consumers like and dislike about you and your product. Ontology data(feedback)—from social media, or any website, or any other source—contains a fortune of helpful business data. In any case, it isn't sufficient to recognize what clients are discussing. You should likewise know how they feel. Sentiment Analysis is one approach to reveal those feelings.

This module uses Naïve Bayes Algorithm to uncover the positive, negative or neutral sentiments in users reviews by comparing them with classifiers which grows on themselves. For an instance, the context should be considered in phrases such as 'This medication along is poison' so here the term 'medication' will be tagged as a negative due to the word 'poison' that is tagged along. In such way, classification of

words in every review is done to bring out positive and negative sentiments on them.

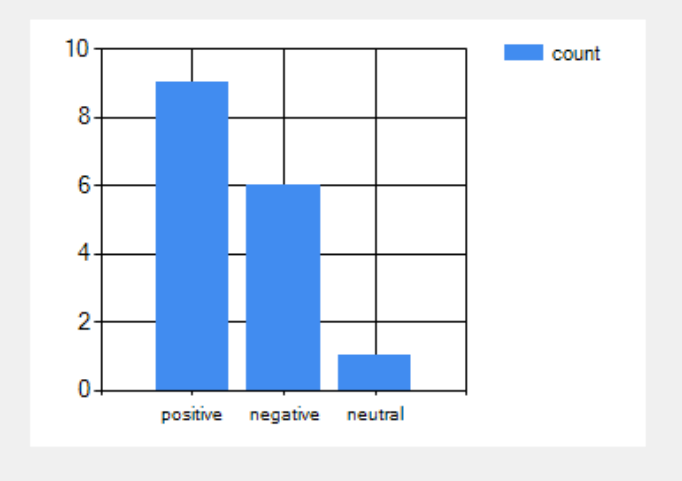


**Figure:** Graphical representation of sentimental analysis of drug

This figures describe the positive, neutral and negative reviews of the patients for a certain drugs using sentimental analysis . We derived the data set from Kaggle.com which included attributes such as patient name, medicine name, reviews and after processing through our software, we obtained 62.5% effectiveness of Ativan (drug) based on the above plotted graph as follows :

Positive-9

Negative-6

Neutral-1

Percentage effectiveness(P.e)=((positive+neutral)/total reviews about medicine)*100

For Ativan,

P.e=((9+1)/16)*100=62.5%

## RESULT

The software converted the feedback data collected from a website in the form of dataset into weighted data to measure consumer thoughts on various drugs such as Ativan, Acetaminophen, Alprazolam, Immune Globulin, and other anxiety and cancer related drugs using positive and negative terms based on their sentiments measured through their comments/reviews. It also suggested whether or not a medicine is good for its prescribed usage based on their percentage effectiveness. Such an approach could be used to raise flags in future clinical surveillance operations, as well as highlighting various other treatment related issues. The majority of the client information was grouped to the territory of the guide connected to positive conclusion, therefore mirroring the general positive perspective of the clients/customers. Ensuing system based demonstrating of the discussion can yield fascinating bits of knowledge on the hidden data trade among clients or consumers. Modules of firmly associating clients can be recognized utilizing a multi scale group discovery technique. By overlaying these modules with content-based data as word-recurrence scores recovered from client posts, we will have the capacity to distinguish data merchants which appear to assume vital parts in the molding the data substance of the gathering. Such merhods could be utilized to bring warnings up in future clinical observation activities, and additionally featuring different other treatment related issues. The outcomes have opened new potential outcomes into creating propelled arrangements, and additionally uncovering challenges in growing such solutions. The agreement on Ativan and different medications relies upon singular patient experience. Web-based social networking, by its temperament, will bring diverse people with various encounters and perspectives.

This arrangement can be imagined on future restorative gadgets that can fill in as post showcasing input circle that shoppers can use to express their fulfillment (or disappointment) straightforwardly to the organization. The organization profits by continuous criticism that would then be able to be utilized to evaluate if there are any issues and quickly address such issues. Online networking can open the entryway for the medicinal services segment in cost lessening, item and administration enhancement, and patient care.

## ACKNOWLEDGEMENT

The Authors are thankful to the authorities of the Department of Computer Science, SRM University.

## CONFLICT OF INTEREST

The Authors have no conflict of interest.

## REFERENCES

[1] Altug Akay, Bj¨orn-Erik Erlandsson, Andrei Dragomir, “Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care” IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICSVOL. 19, NO. 1, 2015.

[2] Jayashri Khairnar,Mayura Kinikar Machine Learning Algorithms for Opinion Mining and Sentiment Classification.

[3] International Journal of Scientific and Research Publications, Volume 3, Issue 6, ISSN 2250-3153

[4] G.Angulakshmi, Dr.R.ManickaChezian, “An Analysis on Opinion Mining: Techniques and Tools”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, 2014.

[5] S. R. Das and M. Y. Chen, “Yahoo! for Amazon: Sentiment extraction from small talk on the Web. Manag. Sci.”, vol. 53, pp. 1375–1388, 2007.

[6] Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes, L. Li, and L. Janacek. "Artificial societies and social simulation using ant colony, particle swarm optimization and cultural algorithms," *New Achievements in Evolutionary Computation*, P. Korosec, Ed. Rijeka, Croatia: InTech, pp. 267–297, 2010.

[7] W. Cornell and W. Cornell. (2013). *How Data Mining Drives Pharma: Information as a Raw Material and Product*. [Online]. Available: http://acswebinars.org/big-data .

[8] L. Dunbrack, "Pharma 2.0 – social media and pharmaceutical sales and marketing," in *Proc. Health Ind. Insights*, 2010, p. 7.

[9] Corley, D. Cook, A. Mikler, and K. Singh, "Text and structural data mining of influenza mentions in web and social media," *Int. J. Environ. Res. Public Health*, vol. 7, pp. 596–615, Feb. 2010.

[10] L. Getoor and C. Diehl, "Link mining: a survey," *SIGKDD Explor. Newsl.*, vol. 7, pp. 3–12, Dec. 2005.

[11] Q. Lu. And and L. Getoor, "Link-based classification," in Proc. 20th Int. Conf. Mach. Learning, Washington, D.C., USA, 2003, pp. 496–503.

[12] Ng, A. Zheng, and M. Jordan, "Stable algorithms for link analysis," in Proc. SIGIR Conf. Inform. Retrieval., New Orleans, LouisianaLO, USA, 2001, pp. 258–266.

[13] Taskar, M. Wong, P. Abbeel, and D. Koller, "Link prediction in relational data," in Proc. Adv. Neural Inform. Process. Syst., Vancouver, B.C. Canada, 2003.

[14] Liben-Nowell and J. M. Kleinberg, "The link prediction problem for social networks," *J. Am. Soc. Inform. Sci. Technol.*, vol. 57, pp. 556–559, May 2007.