Exploring adverse drug reactions of diabetes medicine using social media analytics and interactive visualizations

Pritish Naik
dept. Information Technology
National Institute of Technology, Karnataka
Surathkal, India
pritishnaik.192it010@nitk.edu.in

Rohit Sunil Meshram
dept. Information Technology
National Institute of Technology, Karnataka
Surathkal, India
meshramrohit959@gmail.com

***Abstract*—The aim of this project is to design a framework for an automatically working and live updated social media analysis which also include the interactive data visualization for collecting the information about adverse drug reaction (ADRs) . This framework is developed on the basis of social media data which is now a days one of the detailed data we can found in various subject domain. A diabetes patient data from the online community data like AskaPatient.com is used as a basic and main data for our project. The total number of rating of Glucophage are visualized after the a collection of ADRs posted in Askapatients.com by various users and patients itself. The results are real and real-time and hence it will be helpful for the medical system like the Pharmaceutical companies and the hospitals for tracking the specific medicine and it's adverse drug reaction.**

***Index Terms*—ADRs, FAERS, Glucophage, Askapatient, Social media.**

I. Introduction

The online health community is one of the online platform where the patients can share their experience of the medicine they have taken and its side effect of the medicine. Collecting the data like this is very efficient and accurate way because the data is provided by the patient itself and they also note the side effect of the medication if there any. That's why this data is very important for identifying any adverse drugs reaction of the specific medication as the corresponding information is provided by the many patients itself.

Nearly 60 percent of the scientists and doctors think that the data collected from the social media help to provide the quality services to the patient. On the other hand 41 percent of the users think that the data from social media helps for there choice of the medication and doctors and it is also helpful for them. Due to all these reasons we can conclude that collecting data from the social media is very important and effective for both patients and doctors.

The method of collecting data traditionally is not that useful and efficient as it identifies the data manually.

The goal of the project is that developing a real time and automatic online platform with interactive visualization for analysing the ADRs in efficient way and using this analysis work for providing the better service to the patients and doctors. The main data set for the above framework is taken from the online health community like AskaPatient.com

The issues related to the adverse drug reaction are monitored by the FDA in US in two stages. The first stage is Pre-marketing stage where each drug is tested by medical trial.And the second stage is Post-marketing, in this the hospitals, consumers and pharmaceutical companies are reporting themselves about the issues related to the drug. The main problem with the existing system is that the estimated reporting rate about the ADRs is less than 10 percent and which is not good. And due to this many deaths and hospitalizations takes place which all with combine cost of 75 billion dollars annually. As the traditional statistical methods are insufficient we are using data mining and regression models for discovering the pattern of ADRs and classification and data extraction for obtaining the meaningful information automatically.

The rate of discovering, analyzing and visualizing the OHC data is very less. for doing so we need automatic topic modeling on forum and social media data, automatic information extraction and real time text analysis.

The existing data mining technique is unable to present the result with good visualization, interaction and real time updates.This is the problem why the practitioners are facing difficulties in interpreting the results from the social media for getting the knowledge about the ADRs.

II. Methodology

The whole procedure is divided into 3 sections. they are Data Extraction, Data Analysis, Data Visualization.

A. Data Extraction

To understand the reaction between drugs the author have collected the data from the online health community like AskaPatient.com. In AskaPatient.com the patients share their opinion on specific medication mentioning about its symptoms, diagnosis, drug intake and the data in the from of five point scale rating giving more detailed information. Here for cleaning data we are using python programming language. Now from that we have to extract the data we specifically needed for our use. This new data is now maintained in the csv file as a dataset. diabetes drug (i.e., Glucophage) from the FAERS and and online health community

For collecting the data AskaPatient (http://www.askpatient.com/) is the first choice of the researches. As it provides the side effect of various

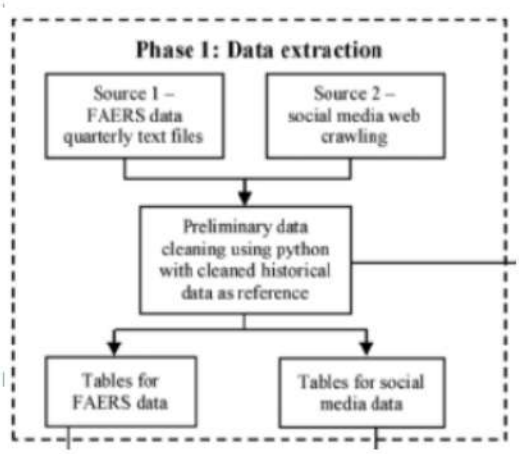

Fig. 1. Data Extraction Phase

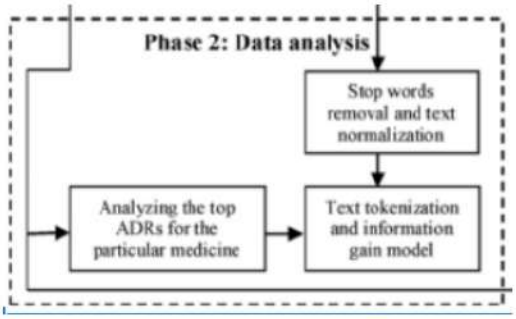

Fig. 2. Data Analysis Phase

medications from year 2000. The data provided by this platform is very specific. For example date.In these dataset the ratings are used to show whether this medicine is useful to the patient, which get scaled from (low; I would not recommend this medicine) to 5 (high; this medicine is very useful to me and I will recommend it). The information provided by this source can be used for recognising the pattern between the variables like age, gender, side effect and rating.

The dataset is stored in a CSV file format for using it in the further phases of the process. Finally we will also get the information about the ADRs of the different drugs and

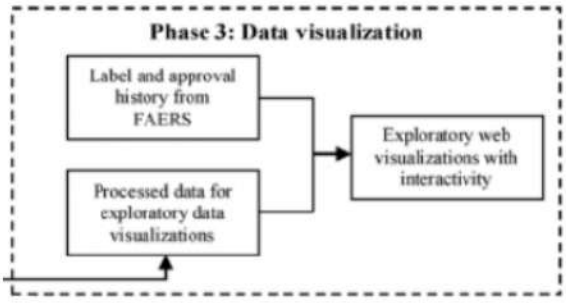

Fig. 3. Data Visualization Phase

medications

B. Data Analysis

In the data analysis we are using tools to read all kind of data like ratings and comments. At this stage we are cleaning the dataset and modeling it for extracting the the ADRs.For that we are using NLP (Natural Language Processing) for processing and manipulating the data. After that python pattern module which is developed by the CLiPS is used here in this phase.

1) Stop words removal and text normalization: data which we have contains some part which does not contain any information. This information is a stop words. here in this phase we remove the stop words which are not needed. Similarly as the data from the Askapatient is in the text form we have to normalize the data in this stage.

2) Extracted Information for ADRs: The third and last step in data analysis is Information Extraction for ADRs. In which based on the information gain each n gram predict the ratings accurately which is given by Askapatient.com. Here rating 1 means very dissatisfied with the medicine and patient will not recommend it. And 5 means the medication is very useful to the patient and patient will recommend it.Now if the side effect is present in the review then it's binary flag in the data is updated as 1, otherwise it will be as 0. in such a way by forming the binary flags for each side effects the presence of there top ADRs are captured in particular reviews.

C. Data Visualization

For the visualization of the dataset we have used the Tableau. This software provides intuitive visualizations for the data with multiple dimensions. The visualization using the Tableau is very interactive and clear for easy understanding.

III. Results

While running the program first we have to import all the dependencies and after that we have to proceed for the program. We have to import the dependencies for the csv file import, pattern3 for the pattern recognition, stopwords for the stopwords recognition and tabulate for representing the data in the tabular format. And some other dependencies will also be there as a supportive for the program.

After satisfying all the dependencies we have to define the function for each and every task which we are going to perform while implementing the idea.

First we have to upload the file which is the dataset file. and after reading the dataset file we have to come across the actual manipulation of the data. The functions will contain the working about normalize the duration of drug consumption, format the gender, format age group etc.

After the programming part we can visualize the data as shown in the figures. Top 3 adverse drug reactions we are visualizing in the graph and we can clearly see in the results that what are the top 3 drug reactions in case of Glucophage medication according to their respective value. top 3 adverse drug reactions are weight change, extreme pain and edema.

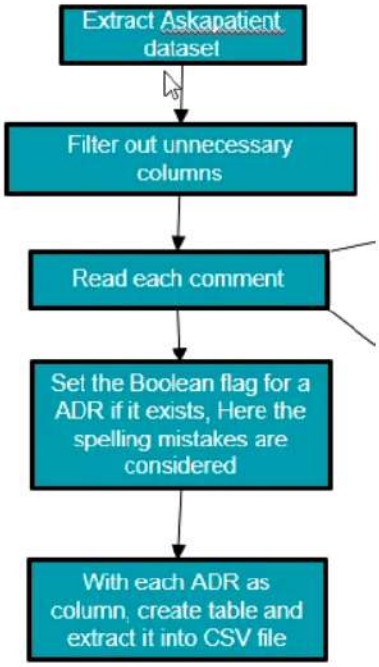

Fig. 4. Flow char for Askapatient.com OHC

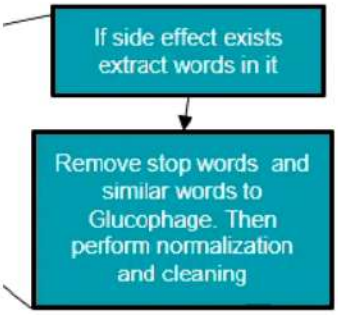

Fig. 5. Normalisation for each comment

Fig. 6.

Similarly we can see top 3 side effect symptoms of the same medication which are Diarrhoea, Stomach issues and nausea. In such a way we have visualized the result of our experiment in the graphical format.

IV. Conclusion

Finally we can conclude that we have enough detailed information about the patient and their interactions about the specific medication thanks to the online platforms like the AskaPatient or the OHC. By using this information we can extract the knowledge we need for the recognition of the ADRs

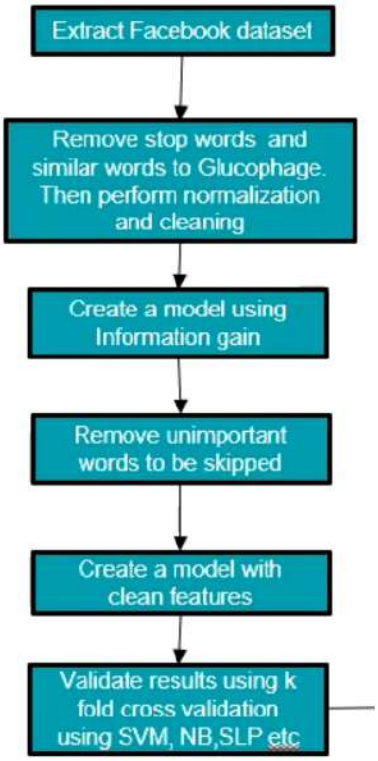

Fig. 7. Flow chart for Facebook

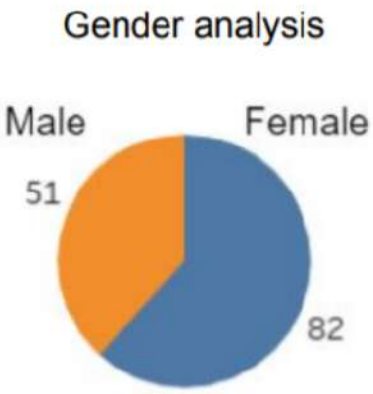

Fig. 8. Gender Analysis

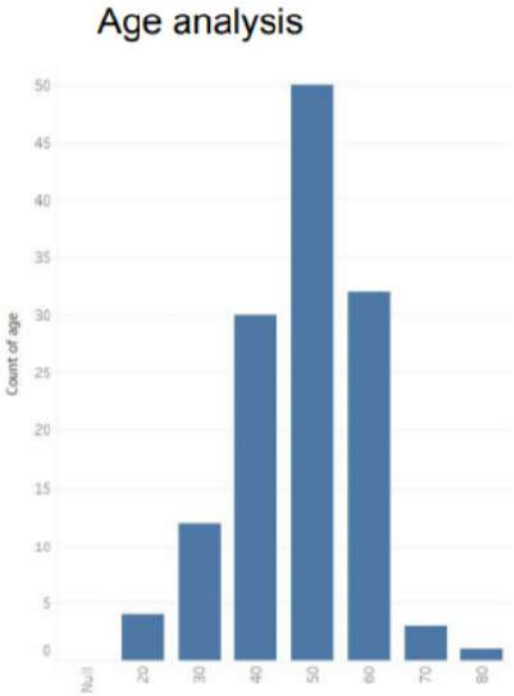

Fig. 9. Age Analysis

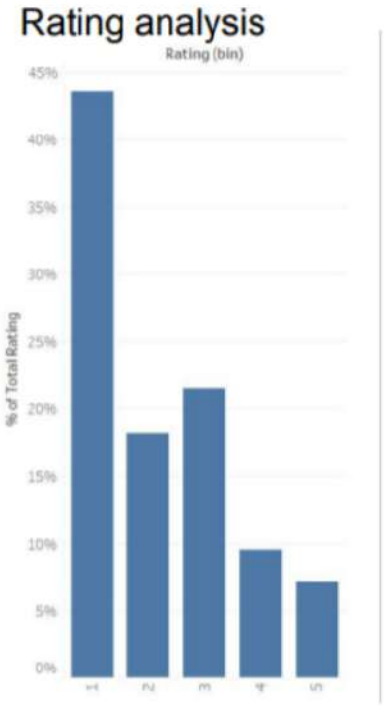

Fig. 10. Rating Analysis

of the specific medication. This Knowledge then can be used for the well being of the patient and the doctors and also for the researchers to do the research work one step further ahead.

We have successfully completed the project by implementing three different phases which is Data Extraction, Data Analysis and Data Visualization by using a data of the dataset from AskaPatient.com. In such a way we can use the social media data or any other online data for the meaningful purpose after gaining the information and transforming the information according to necessity.

ACKNOWLEDGMENT

We are very thankful to Dr. Sowmya Kamath S. for their valuable guidance to us while our project work named "Exploring adverse drug reactions of diabetes medicine using social media analytics and interactive visualizations". Which doing this project we came across many difficulties but it was the phase from which we have to go for learning further ahead by facing these difficulties. We are really very much thankful to them for their support.

REFERENCES

[1] Akay, A., Dragomir, A., Erlandsson, B. E. (2015). Network-based modeling and intelligent data mining of social media for improving care.*IEEE Journal of Biomedical and Health Informatics*, 19(1), 210–218.

[2] Almansoori, W., Addam, O., Zarour, O., Elzohbi, M., Sarhan, A., Kaya, M., ... Alhajj, R. (2014). The power of social network construction and analysis for knowledge discovery in the medical referral process. Journal of Organizational Computing and Electronic Commerce, 24(2-3), 186–21.

[3] Aswani, R., Kar, A. K., Ilavarasan, P. V., Dwivedi, Y. K. (2018). Search engine marketing is not all gold: insights from Twitter and SEOClerks.*International Journal of Information Management* , 38(1), 107–116.

[4] Bate, A., Evans, S. J. W. (2009). Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiology Drug Safety, 18, 427–436.

[5] Benton, A., Ungar, L., Hill, S., Hennessy, S., Mao, J., Chung, A., ... Holmes, J. H. (2011). Identifying potential adverse effects using the web: A new approach to medical hypothesis generation. *Journal of Biomedical Informatics*, 44(6), 989–996.

[6] Bian, J., Topaloglu, U., Yu, F. (2012). Towards large-scale twitter mining for drugrelated adverse events.*Proceedings of the 2012 International workshop on smart health and wellbeing*, 25–32.

[7] Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., Ebert, D. S. (2014). Public behavior response analysis in disaster events utilizing visual analytics of microblog data. Computers Graphics, 38, 51–60.

[8] Chau, M., Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. MIS Quarterly, 36(4), 1189–1216.

[9] Chee, B. W., Berlin, R., Schatz, B. (2011). Predicting adverse drug events from personal health messages.*In AMIA Annual Symposium Proceedings*, Vol. 2011, 217

[10] Chen, H., Chiang, R. H., Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165–1188.

[11] Chretien, K. C., Kind, T. (2013). Social media and clinical care ethical, professional, and social implications. Circulation, 127, 1413–1421.

[12] Cocos, A., Fiks, A. G., Masino, A. J. (2017). Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *Journal of the American Medical Informatics Association*, 24(4), 813–821.

Rohit and Pritish

ORIGINALITY REPORT

14%	4%	11%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Si Li, Chia-Hui Yu, Yichuan Wang, Yedurag Babu. "Exploring adverse drug reactions of diabetes medicine using social media analytics and interactive visualizations", International Journal of Information Management, 2019 Publication	7%
2	dblp.uni-trier.de Internet Source	2%
3	ijcnwc.org Internet Source	1%
4	Submitted to Rochester Institute of Technology Student Paper	1%
5	www.ijcaonline.org Internet Source	1%
6	Nick Hajli, Michel Laroche. "Applications of business intelligence and analytics in social media marketing", International Journal of Information Management, 2019 Publication	1%

7	www.arpnjournals.org Internet Source	<1%
8	"Smart Health", Springer Science and Business Media LLC, 2017 Publication	<1%
9	Wadhah Almansoori, Omar Addam, Omar Zarour, Mohamad Elzohbi, Abdullah Sarhan, Mehmet Kaya, Jon Rokne, Reda Alhajj. "The Power of Social Network Construction and Analysis for Knowledge Discovery in the Medical Referral Process", Journal of Organizational Computing and Electronic Commerce, 2014 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		