Disease Analysis of the Required Zone Using Big Data Analytics

Aniruddha Prabhu B P¹, Likith Reddy S², Nandeesh Changavar Matt³, Manoj K L⁴

¹,²,³,⁴ Department of Computer Science and Engineering, Siddaganga Institute of Technology, Tumkur, India
Email: aniprabhubp@gmail.com¹, reddy.likhithreddy@gmail.com², nandeeshcm7@gmail.com³, manojvedhant.1997@gmail.com⁴

Abstract- Nowadays there is huge growth in big data analytics in every field. Huge amount of data have been generating in every field. Especially in medical field, huge data is being generated every day. This data can be used for analytics for various purposes. With the data generated from healthcare communities and other medical fields, we can make accurate analysis to detect diseases early, can take precautions, and can find the cause of the disease. In this paper we are analyzing the data which we collected from various medical fields across certain area. We analyze the data and predict which disease is more likely to occur further in which particular place. So that, it may be helpful for the people in that particular place can take precautions of that particular disease.

Keywords: Big Data, HealthCare, Data Predictive Analysis.

1. INTRODUCTION

Big data is described as 3 V's: Huge volume of data, wide variety of data and the velocity or the speed at which data is processed. Usually, the size of the data considered for analysis has no limit. [6]. Big data analysis processes large amount of data and uncover the hidden patterns, correlations and other insights. With modern technology, it is possible to analyze the data and get results from them. Big data analysis is a task that is slower and less efficient with lots of traditional business intelligence solutions.[7][8]

Predictive analysis is the process of extracting useful information from the existing data sets in order to find patterns and predict future outcomes and trends. Predictive analysis will not tell us what is going to happen in the future. It tells us what might happen in the future. Predictive models and analysis are normally used to predict future chances. In case of business, predictive models are used to analyze the collected data and facts to understand customer's behavioral needs and to identify all the available opportunities for a company. It uses different techniques, for example data mining, statistical modeling. It will help analysts to make future better. As the same case applies to medical field where the analysts analyze the data to understand about the symptoms of the disease, nature in which the diseases is occurring.

It will help to avoid the diseases, and to take precautions against those diseases. In this paper, we are collecting the huge amount of data generated in the medical field and subject it to the analysis to find the hidden patterns and reasons for the diseases and at which place the particular disease is more likely to occur. By the analysis done on the data sets, we can predict future outcome like the type of disease which might occur at a particular place. So the reason for the disease can be found and the necessary precautions can be taken to overcome the disease.[4][5].

2. LITERATURE SURVEY

2.1 The Twitter Bable

In this paper author shows how the data can be gathered and assembled from the mobile devices, and also urges that the data so obtained is more accurate than the data obtained from the surveys. The classical data sources such as surveys; census will not have higher geographical resolution and are restricted to the weekends and weekdays. The data coming from the mobile devices can be easily located with respect to the date, time and location.

The study is focused on the areas of the Barcelona and Madrid. Where the three different types of data are available, and also they have assessed the association of the different aspects of the data they are temporal evolution of the people, the spatial scattering of the people, mobility patterns of the people. The result shows that the three aspects will provide the comparable information. The representativeness of the twitter data is lower than the mobile data and the census, survey data.

Author also suggests using the social media data for the flu surveillance because they have the potential to cover the large people. The social media data declines the
media attention because the social media data will increase the chatter messages; these messages are related to the flu but will not give the exact information about the types of the flu that the people are suffering from. This paper summarizes the algorithm that is developed for influenza infection detection.

It also categorizes the relevant tweets from the other chatter. The US health care system is adopting electronic health records for maintaining the information about the patients, which will also increase the total amount of the clinical data which are required for the analysis of the disease. As a result of this, rapid progress has been made in the field of the clinical analytics. Techniques for analyzing large quantities of data and gaining new insights from that analysis which is part of what is known as big data [1].

The data is increasing very vastly day by day. This data is very much useful in many fields. The huge volume of data is often termed as big data. In the healthcare industries also the data which is getting generated is huge. This data can be used by different people across different fields in the healthcare. The data can be used in the analysis of the disease, in predicting the disease, its behavior and other things. The author also insists upon the hurdles or challenges that arise in the usage of big data in healthcare. Some of the technical challenges are lack of established standards for healthcare data, different data generated from different models because of old systems and modernized systems, issues with the hospitals for collecting the actual data and systems which are not compatible with the advanced technologies, inadequate real time processing, lack of quality in the data, protecting the privacy of the patients and many more.

Finally the author comes to a conclusion that in recent times the people are conscious about the health and are ready to spend more amount and the healthcare industries are providing better healthcare facilities in the interest of the people. Big data is vastly used in healthcare industry and researchers are using big data for health data analysis, eHealth services, to develop systems for the diagnosis of the diseases, to provide privacy for personalized treatment and to prevent/control chronic diseases [2][3].

3. PROPOSED SYSTEM

In this paper we are going to make use of additional information related to the areas/regions along with the healthcare data. The information regarding the disease, symptoms and other prescription data of a patient will be available in the patients’ dataset. Using these data we are analyzing and trying to bring out similarities among the patients and their behaviors. The disease similarities information can be gained with some transformations using map reduce. Here we are proposing a solution to find the similarities in the patients’ disease data based on their location. These locations data can be used for further analysis by various departments of the government related to healthcare as well as by the hospitals.

3.1 Step by step process

i. Collecting the healthcare data by the data providing company. Usually the companies will have their representatives who will be assigned the task of collecting the data from the different sources by paying some money.

ii. Once the data is collected, the data should be cleansed and processed to obtain the correct data so that the data which we will be using should have some value.

iii. After the data is processed the big data analyst will analyze the data which is there and will take the data fields which are necessary and useful for the further processes.

iv. After the analysis is done, the analyst will extract the data from the necessary useful data fields.

v. By using this data, the analyst will obtain the information which will be useful for further processing of the data and to carry out some effective decisions by the organizations related to healthcare.

We are collecting the healthcare data and performing some operations on it. The healthcare data can be used for many purposes such as disease prediction, fraud prevention, pharmaceutical market trends, etc. Here we are predicting the geography locations where a particular disease rate is significantly high. Using this analysis the government can take some countermeasures for the particular diseases for preventing or curing the disease. The healthcare data which is obtained will be divided based on the properties/fields.

The properties such as diseases and the regions will be taken for this analysis. The diseases which are there in a particular region are taken and the count of the persons who are affected by the disease is obtained and using which the regions will be checked if there is any abnormal
trend in the count of a particular disease. This will be useful for the health organizations such as hospitals, pharmacies and others to gain some information about the disease trends in a particular region and can take further actions accordingly.

Hospitals may take some decisions in preventing the particular disease, diagnosing the disease as well as predicting the disease based on the symptoms which were found for a disease based on the previous patients data. Pharmaceutical companies can increase the medicines and drugs supply related to that particular disease based on their estimation of number of persons who might get the same disease and in turn can increase their product sales. Government can use this information and take several precautions and prevention methods to make the life in that particular region little affected. Figure 1 shows the use case diagram of the proposed solution. Here data represents the healthcare data.

4. RESULTS

As seen in the graph below, it can be seen that in the y-axis the data is the approximate count of the people who are affected with a particular disease. The graph will show the regions along the x-axis. Overall the graph gives some facts about the affection of some of the diseases in a particular region. This can be helpful for different healthcare organizations to take some preventive measures and get some medical assistance in abundance for that particular region where a disease in abnormally high than the others. This can be helpful for the government to alert the people of that region to take care of themselves against that disease and can guide them with some preventive measures not to get affected by the disease or can even send some medical assistance to that particular region.

![Fig. 1 Use case diagram of entire process](image)

5. CONCLUSION

In this paper we have proposed a new idea of fetching the useful information, from the people regarding the healthcare. This information can be used for different purpose by the peoples. Number of techniques to study trends of disease causing intensity in a particular region is very less in number. With the help of the techniques proposed in our paper analysis can be made to predict the cause of the disease and similarities in the affected people so as to take decision for reducing the disease causing intensity.

The proposed technique can be further improvised and can be used by the government to carry out the advanced researches in the health care field and to take the precautions regarding the diseases which are likely to occur. Further analysis can be made to classify the diseases based on different aspects such as occurrence of the disease in a particular gender, Weather conditions and geography, etc.
REFERENCES


[6]. The Twitter of Babel: Mapping World Languages through Micro blogging Platforms Delia Mocanu, Published: April 18, 2013.
