Analysis on Drug based on Patient Reviews using Big Data

Priyanka S  
Department of Information Technology  
Jerusalem College of Engineering, Pallikaranai, Chennai

Supriya M  
Department of Information Technology  
Jerusalem College of Engineering, Pallikaranai, Chennai

Lilly Sheeba S  
Department of Information Technology  
Jerusalem College of Engineering, Pallikaranai, Chennai

Abstract

Big data analytics deals with large number of data sets with volume and velocity beyond the ability of typical database software to store, manage, capture and analyze. Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making. This result in greater operational efficiencies, cost reductions, and reduced risks for the business. In existing system it is more difficult to handle more data and to predict reaction. In proposed system, useful information is taken from the feedback and comments. A summarized solution is created to carry out analysis to specify appropriate drugs for categorized patients.

Keywords: Big data, forum, drugs, patient reviews, user polarity, review pertinence

I. INTRODUCTION

Big data overshoots the processing capacity of typical database systems. The data is either too big or moves too quick, or does not lay the structures of database architectures. To gain value from this data, you must choose an alternative way to process it. Database management system and desktop statistics and visualization packages often have difficulty in handling big data. The work instead requires massively parallel software running on tens, hundreds, or even thousands of servers. Big data varies depending on the capability of users and their tools, and expanding capabilities to make big data a moving target. Some organizations, facing hundreds of gigabytes of data for the first time may activate a necessity to get data management options. For others, it may take tens or hundreds of terabytes before data size becomes a notable consideration.

Characteristics of Big data include volume where the quantity of data that is generated is very important. It is the size of the data which determines value and potentials of the data under consideration. The other aspect of big data is Variety which means that the category to which Big Data belongs to is a very essential fact that needs to be known by the data analysts. This helps people, who are closely analyzing the data and are associated with it, to effectively use the data to their advantage and thus upholding the importance of the Big Data. Big data also deals with Velocity it refers to the speed of generation of data or how fast the data are generated and processed to meet the demands and the challenges which are in the path of growth and development. Variability is handled by big data this is a factor which can be a problem for those who are analyze the data. This refers to the incompatibility which can be shown by the data at times, thus hampering the process that can handle and manage the data effectively. Data management can become a very complex process, especially large volumes of data come from different sources. These data need to be linked, connected and correlated in order to grasp the information that is supposed to be conveyed by these data. This situation is therefore, termed as the complexity of Big Data.

II. RELATED WORK

Big data define different types of data sets consists of unstructured and structured data. Data sets are stored in databases that become more complicated and extraordinary to store, manage and analyze via typical database tools. The huge increase in the data sets generated on the digital universe is increasing exponentially. By using current technologies to store and analyze, the massive volume of data are not up to the level, since which is unable to extract the required sample datasets. Therefore, they design an architectural platform to analyze both the remote access real time data and offline data.

Big data analytics help us to make better decisions. Therefore by using Big data, modifications in paradigms are analyzed. In healthcare scenarios, medical peoples gather large volume of data about medical history, patient history and patient details. The nature of that data are very complex, and sometimes the practitioners are not able to show a relationship with other information, which result in missing of important information.
III. PROPOSED SYSTEM

In proposed system the useful information are taken from the feedback and comments. By enhanced review based algorithm the data are analyzed. Hadoop tool is used for storage and processing. In this system it consists of three modules. Module 1 deals with dataset collection from multiple social platforms. Based on these collections spam and non-spam reviews are identified which helps to identifying review pertinence and in eliminating biased entries. This work is carried out in three phases. The first phase is about the detection of user polarity. In this work user feedbacks are analyzed which may be in long textual forms from which useful data are taken from comments. A new library is constructed based on user sentiments in medical reviews Words are matched with the corresponding words in mesh library. The average calculations are performed to determine user polarity. Last phase deals with enhanced review based algorithm is used which can effectively handle opinions on diseases. An Integrated User Interface is created to handle all operations into one system. This algorithm mainly used to classify drugs based positive and negative reviews which precisely pinpoints the reasons for user satisfaction/dissatisfaction on the drugs. Alternate drugs/suppression of side effects can be triggered and summarized solution is created. Fig. 1 gives the architecture diagram employed in our proposed system.

IV. CONCLUSION

In the proposed system feedbacks are classified according to the enhanced review based algorithm. Multiple data sets are handled by the tool Hadoop using map reduce algorithm. Alternate drugs/suppression of side effects can be triggered and summarized solution is created. Big data include innovation, growth and long term sustainability. Big data is exploited in an open and transparent manner.

REFERENCES


