Efficient Prediction of Medicine usage Using ML by considering Social Media Dataset

Sindhu S.B[1], Prof. Dr. Verappa B.N[2]
[1,2]*Department of studies in computer science & Engineering, University BDT College of Engineering (A constituent college of VTU, Belgavi), Davangere, Karnataka-577001, India*

***Abstract*- This paper based on the research carried out in the area of data mining depends for managing bulk amount of data with mining in social media on using composite applications for performing more sophisticated analysis. Enhancement of social media may address this need. The objective of this paper is to introduce such type of tool which used in social network to characterised Medicine Usage. This paper outlined a structured approach to analyse social media in order to capture emerging trends in medicine abuse by applying powerful methods like Machine Learning. This paper describes how to fetch important data for analysis from social network. Then big data techniques to extract useful content for analysis are discussed.**

Index terms- social media; data mining; Big data.

INTRODUCTION

Social network (media) is one to extracting the information from the internet. Nowadays it is used for extracting the data of patient's to know the understanding of patient symptoms. Social media, classify from individual messaging to live for as, is providing immeasurable opportunities for patient to Converse their experiences with drug and devices. Social media allows message contribution, gathering information and distribution in the health care space. Health care is one which contains the information of patients with their permission. It provides an effective social networking environment. The proper way of mining information and drift from the knowledge is cloud. Using network based analysis method it model the social media such as Facebook, Twitter, WebMD. Nowadays the scientific experiment often requires bulk amount of computation during simulation and data processing. Performance of super computer is increasing rapidly. It allows solving scientific problem by automatic computational through collection or array list which is emerged by set of sensors. Nowadays electronic mechanism is growing in recent scenario.

Numerous studies have been published recently in this realm, including studies on pharmacovigilance,2 identifying smoking cessation patterns,3 identifying user social circles with common experiences (like drug abuse),4 monitoring malpractice,5 and tracking infectious disease spread.6–8 A systematic review9 conducted in 2014 found numerous attempts to use this user-generated data, but none yet integrated in national surveillance programs, noting the promise and challenges of the field quite succinctly: "More direct access to such [social media] data could enable surveillance epidemiologists to detect potential public health threats such as rare, new diseases or early-level warnings for epidemics. But how useful are data from social media and the Internet, and what is the potential to enhance surveillance? The challenges of using these emerging surveillance systems for infectious disease epidemiology, including the specific resources needed, technical requirements, and acceptability to public health practitioners and policymakers, have wide-reaching implications for public health surveillance in the 21st century."9 The use of social media for health monitoring and surveillance indeed has many drawbacks and difficulties, particularly if done automatically. For example, traditional NLP methods that are applied to longer texts have proven to be inadequate when applied to short texts, such as those found in Twitter.2 Something seemingly simple, such as searching and collecting relevant postings, has also proven to be quite challenging, given the amount of data and the diverse styles and wording used by people to refer to the topic of interest in colloquial terms (semantic heterogeneity) inherent to this type of media. The goal of this session was to attract researchers that have explored automatic methods for

the collection, extraction, representation, analysis, and validation of social media data for public health surveillance and monitoring, including epidemiological and behavioral studies. It serves as a unique forum to discuss novel approaches to text and data mining methods that respond to the specific requirements of social media and that can prove invaluable for public health surveillance. Research topics presented at this session include: • Early detection of disease outbreaks • Medication safety, including medicine interactions and dietary supplement safety • Health behaviors, including diet success • Individual well-being which affects mental and physical health.

LITERATURE SURVEY

Noemie Elhadad, et al[1], agreeable chains are a noteworthy hotspot for customer created criticism on about all items and administrations. Clients much of the time accept on social bind to reveal now and then genuine episodes as opposed to going to social correspondence channels. This vital, significant, customer made certainties, if extricated genuinely and powerfully from the social chain, can possibly have the positive effect on basic applications identified with social wellbeing and security, and past. Shockingly, the creation of data from social chain where the yield of the extraction procedure is utilized to take solid activities in the genuine world are not very much upheld by existing innovation. Customary data creation approaches don't function admirably finished the exceedingly casual and ungrammatical sentence structure in social chain. They don't deal with the generation and collection of uncommon substance. In our progressing aggregate undertaking between Columbia University and the New York City Department of Health and Mental Hygiene (DOHMH), this paper intend to address these distinction in research and innovation for one essential general wellbeing.

Erwan Le Martelot et al[2] Today wherever organize is accessible. The people group exposure got an expanding consideration as an approach to uncover the arrangement of systems and associated inside than externally. However the vast majority of the powerful techniques accessible don't think about the conceivable levels of association, or scales, a system may incorporate and are in this manner restricted. In this paper Author said in regards to perfect with worldwide and neighborhood criteria that empowers quick multi-scale group finding. The strategy is to clarify with two calculations, one for each sort of measure, and executed with 6 known standard. Disclosure people group at different level is a computationally extravagant assignment. Consequently, this activity puts a solid consideration on the lessening of computational unpredictability. A few heuristics are initiated for accelerate reason. Trial display the competency and correct of our way regarding singular calculation and model by testing them against substantial out-comes in multi-scale arrange. This work likewise offers an appraisal amongst criteria and between the worldwide and nearby methodologies.

Big information is the term that portrayed by its expanding volume, speed, assortment and veracity. Every one of these qualities makes handling on this huge information an unpredictable undertaking. Along these lines, for preparing such information Author need to do it any other way like Map Reduce Framework. At the point when an association trades information for mining helpful data from this Big Data then protection of the information turns into an imperative issue in the earlier years, a few security safeguarding models have been given. Anonymizing

[3] The dataset should be possible on numerous operations like speculation, concealment and specialization. These calculations are for the most part reasonable for dataset that does not have the attributes of the Big Data. To propagate the protection of dataset a calculation was proposed recently. An creator speaks to how the development of enormous Data qualities, Map Reduce structure for security safeguarding in eventual fate of our exploration.

E. Srimathi, K. A. Apoorva[4], as of late numerous web administrations expect customers to share their private electronic wellbeing records for inquire about examination or information mining, which prompts security issues. The size of information in cloud foundation ascends as far as nature of Big Data; in this manner making it a contention for conventional programming instruments to process such mass information inside a tolerable slipped by time. As a consequence, it is a contention for current anonymization strategies to save protection on classified extensible informational collections

because of their deficiency of scalability. An Author speaks to an extensible two-stage way to deal with Anonym zing versatile datasets utilizing dynamic Map Reduce structure and LKC security display.

METHODOLOGY

The fig 1 below is the proposed architecture which is used to mine the social media data. The diagram consists of database where the social media data is stored and from that the particular data is selected and extraction of the data is done. Data extraction is nothing but exacting a particular data of feature is selected.

After this step data transformation is done, during this phase transforming a particular data is done. And those data is the send to process by the machine learning model where those data is tested in which class they belong. There is a training set of instances present in the machine learning model which is used to predict the new test instances where or which class they belong.

The algorithm works as listed steps the first step is to collect the raw data from the social media as the input.

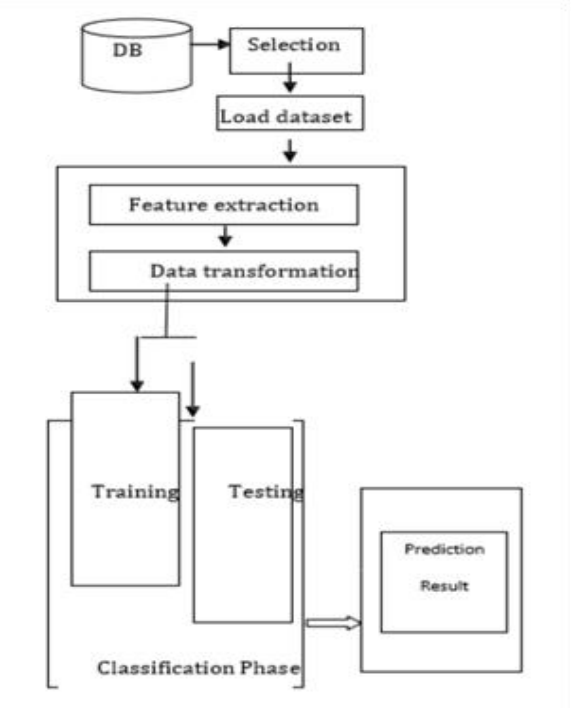

Fig: Block Diagram

The next step after the input is to apply supervised learning algorithm for feature extraction. And at this particular step the output is also generated as the extracted data. After this step the training set is generated and even the testing data is created by applying various algorithms. Next step is to apply machine learning algorithm to the training dataset to train the model for getting the correct predicted output. After this step are main aim is to build the classifier model to classify the test data. Its upto us that which classifier algorithm we can use. There are many classifier algorithm, they are listed above in this paper. After building the classifier model the next step is to apply that model for classifying the training data sets. The last step is we get the output as the predicted class of that particular training data se

IMPLEMENTATION

Prediction of Medicine Drug usage to identify requirement, availability and scarcity of medicine in different places based on the usage of medicine in different areas of cities. For the prediction of medicine usage we are developed a machine learning software model to which we are providing dataset information which contains information regarding which medicine is used by which patient and in which area along with additional properties. After provide dataset which is unreadable format it need to convert to structure one for further processing. After conversion we are going to extract features of medicine dataset which are necessary for identifying their usage using supervised learning technique. Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object and a desired output value. A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. After classifying data labels they are going to group which are related to each other based on the subject which we are considered. Then cluster is going to form the group of similar elements like on medicine name, areas etc., supervised

learning method categorizes our data into a desired and distinct number of classes where we can assign label to each class. Here are used decision tree classification technique to make decision on available data items and classify them according to critical information. Decision Tree is simple to understand and visualise, requires little data preparation, and can handle both numerical and categorical data.

RESULT

Result on the studies is carried out with large number of data set collected from real world health organization. Classification is done on the dataset with feature extraction resulting in different categories of data is available on different medicines. Here we are going to analyze data based on usage like medicine name, area, gender like this we are considered a dataset processed according requirements and resulted in 98 percent accuracy using machine learning technique

CONCLUSION

This paper presented our approach for mining and managing data from social chain which depends upon combination of bulk amount of data from social networks which is based on combination of big data and infrastructure paradigms. Machine Learning model is used to mine, store and process bulk data through social network. Processing of mined data is also performed by ML which simplifies development of new algorithms and provides high scalability and flexibility. This paper presents development of an implementation of Machine Learning Technique that extend to large chunks(storage) of machines comprising thousands of machines. The utilization makes efficient use of these machine resources is suitable for many large computational issue encountered at Google.

REFERENCES

[1] Noemie Elhadad, et al “Information extraction from social media for public health”.
[2] Erwan Le Martelot, “Fast multi scale detection of relevant communities”.
[3] Hari Kumar.R M.E (CSE), Dr. P. Uma Maheshwari , Ph.d, “Literature survey on big data in cloud,” International Journal of Technical Research and Applications e-ISSN: 2320-8163.
[4] E.Srimathi, K.A. Apoorva “Preserving identity privacy of healthcare records in big data publishing using dynamic MR”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5, Issue 4, 2015.