

**SHORT COMMUNICATION****Hadoop Based Biomedical Document Analysis**

Parameswara Reddy A\*

Vignan Pharmacy College, Guntur, India

\*Correspondence to: Parameswara Reddy A, Email: parameswar.reddy864@gmail.com

Received Date: 23 June 2020; Accepted Date: 22 July 2020; Published Date: 28 July 2020

© Copyright: Parameswara Reddy A. First Published by Allied Academies. This is an open access article, published under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>). This license permits non-commercial use, distribution and reproduction of the article, provided the original work is appropriately acknowledged with correct citation details.

In the biomedical research collecting and searching the publications or documents plays a key role, because the publications are unstructured data not grouped according to the keywords. From the few decades the publications has been increased exponentially in the field of bioinformatics, so it is a difficult task for a user to search the relevant data based on the user criteria for decision making. In this paper we discuss about traditional data mining extraction to latest document extraction and analysis. In the bioinformatics an ecosystem that transforms case-based studies to large-scale, data-driven research in big data.

The characteristics of big data are defined by 3Vs: volume, variety, and velocity. The challenges of bioinformatics are storing, managing, and analyzing massive amounts of medical datasets. The big data require novel technologies to extract relevant documents and enable health-care solutions. In large datasets multiple technologies and multiple data sources are used together, such as Artificial intelligence (AI), Machine Learning, Hadoop and Data mining tools. The automatic classification of medical documents into predefined classes is growing rapidly on online data repositories, one of the biggest problems motivated to assist experts in finding useful information from a large amount of distributed document repositories. In distributed biomedical systems, text classification models are important as it can lead to advances in decision making including gene functions, gene-disease patterns, gene-gene associations and Medical Subject Heading (MeSH) knowledge discovery. It is important to classify and organize the biomedical databases so users can access the useful information easily and quickly. Computerizing wellbeing observing favors a proactive methodology that alleviates medicinal facilities by sparing costs identified with hospitalization, and it likewise upgrades human services benefits by improving sitting tight time for counsels. As of late, the quantity of information sources in social insurance industry has developed quickly because of broad utilization of portable and wearable sensors innovations, which has overwhelmed human services territory with a tremendous measure of information. Hence, it winds up testing to perform

medicinal services information examination dependent on conventional strategies which are unfit to deal with the high volume of enhanced medical information. In general, healthcare domain has four categories of analytics: descriptive, diagnostic, predictive, and prescriptive analytics; a brief description of each one of them is given below.

**Descriptive Analytics:** It consists of describing current situations and reporting on them. Several techniques are employed to perform this level of analytics. For instance, descriptive statistics tools like histograms and charts are among the techniques used in descriptive analytics. **Diagnostic Analysis:** It aims to explain why certain events occurred and what the factors that triggered them are. For example, diagnostic analysis attempts to understand the reasons behind the regular readmission of some patients by using several methods such as clustering and decision trees. **Predictive Analytics:** It reflects the ability to predict future events; it also helps in identifying trends and determining probabilities of uncertain outcomes. An illustration of its role is to predict whether a patient can get complications or not. Predictive models are often built using machine learning techniques. **Prescriptive Analytics:** Its goal is to propose suitable actions leading to optimal decision-making. For instance, prescriptive analysis may suggest rejecting a given treatment in the case of a harming side effect high probability. Decision trees and Monte Carlo simulation are examples of methods applied to perform prescriptive analytics.

In biomedical research, big data frequently contains an assortment of datasets from different information sources like Medline/Pubmed, Epigenomics, PROMIS, EyeGENE etc including authorized, randomized or non-randomized clinical investigations, distributed or unpublished information, and medicinal services databases. Thus, it is a concern whether the enormous data is a genuinely illustrative of the target patient populace with the maladies under investigation since conceivable determination predisposition may have happened while tolerating individual data indexes into the huge data. Here the search results are shown in reverse chronological order, i.e., it shows documents according to the date and time or frequently

accessed list. Boolean operators were integrated along with MeSH (Medical Subject Heading) terms for document retrieval according to the query construction, MeSH deals with the real content of articles. MeSH database is responsible for finding

and choosing MeSH terms, check the definition and document entity information make PubMed search strategy, show MeSH hierarchy, associate sub-headings and establish a link to MeSH browser.