

A Survey on Skin Disease Data Analysis Using Big Data Analytics

Sanjeev Kumar
School of Computer Science
Lovely Professional University
Phagwara, Punjab

Abstract

Data analytics is known as a method of cleaning, transforming, modeling and analyzing the large set of data to discover meaningful information and to help in the process of decision making. Data analytics consist of multiple facts and methodologies. Analysis of skin disease is being done utilizing datasets with big data analytics and systems such as Hadoop. Datasets consist of images or text formats and other sources of information. The analysis is done to gain information and insight about the disease which affects the human. Various methods and techniques are being utilized for this analysis purpose. These could be classified as per the properties and attributes.,

1. Introduction

Data analytics is the science of analyzing raw data in order to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption. Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a business or system.

2. Data Analytics

It aims at obtaining raw data and transforming it into useful information. This information is then used in decision making and to test the hypothesis and disprove theory. It is a procedure to analyze the data and a technique for interpreting the results. There are few terms in the data analytics.

a. Data collection

In this phase, raw data is collected from different sources. It may also be collected from sensors which are monitoring natural resources, some real life environments such as traffic cameras, satellites, recording devices, and healthcare. Third party data sources could also be used for rare or complex data.

b. Data processing

Data obtained from the data source must be processed and organized in the required format for the analysis. Data could be of any format: structured, semi structured or unstructured. In data processing, the data must be uniform for the next phase.

c. Data cleaning

This is an important process of data analytics. Anything wrong happen in this phase could degrade the quality of whole performance and the integrity of data may be challenged. In this step, the input is the real data set collected from the real world data which is new. There are chances of inaccurate data, duplicate data, formatting issue, un-structured data, redundant records, naming conventions and incomplete data. The increase in such data leads towards incorrect and non-desirable results. In this process our primary goal is to remove redundancy and incomplete data from the data base. This is because when we apply any query on the database than that time it takes a lot of time to result an output. So to increase the probability of its correctness we need to remove its incompleteness and redundancy. This is achieved by applying various algorithms, tools and techniques which complies a complete and unique data in the database. And then when we apply query on that unique filtered data, we get correct output as a result which will be used for taking various decisions for the organization and various predictions in-favor of the organization.

3. Skin Disease Analysis

Human skin creates the partition between the inside of our body and the outside. It guards the inner organs from harmful bacteria and viruses, and also regulates our body temperature. Some skin diseases are acne, hives, allergies, melanoma, and skin cancer. In skin disease analysis, we use the dataset of particular disease and analyze that. For this purpose, a quantitative and qualitative analysis technique will be used, such as how many people are suffering from the particular disease etc. Some analysis techniques are as follows:

a. Data Mining Techniques

Data mining is a procedure within which the growth is spelled out by finding trends or patterns in a dataset. It is the most beneficial technique in a tentative analysis scenario where there are no prearranged concepts about the conclusion which will be derived. The knowledge results

obtained from data mining processes are used to assist human decision makers and finally to solve their complex problems.

b. Prediction

This method of data analysis evaluates the numeric and well-ordered future values grounded on patterns of a dataset. Prediction attributes are always continuous and not categorical. Predictive analysis is a statistical method which takes data as input and extracts meaningful information from the dataset. The extracted information is then used to predict behavioral trends and patterns. In simpler terms, Predictive analysis tends to learn from the dataset to predict the future trend for the sake of better decision making.

c. Regression

This method discloses the relationship between dependent variable and independent variable. It describes the change in the value of a dependent variable when any one of the independent variable is changed while others remain constant. Regression technique follows mathematical statistical approach. Other mathematical methods used under regression are linear regression and logistic regression.

d. Visualization

Visualization is termed as the modest and easier way to epitomize data in an understandable data presentation. The input is multifaceted data and subsequent easy and clear uncovered patterns of complex data that can be effortlessly recited by users through the progression of data mining.

This is closely related to the pattern detection capabilities of a human. This is done with the help of certain factors like size, vision and previous experiences.

4. Attributes of Big data

Big data is the term for a collection of data sets for large no data and complex that it becomes difficult to process it using traditional database management tools or data processing application. In big data, three terms: volume, velocity and variety, are frequently used to describe different aspects of the big data (refer figure 1). These attributes define the nature of the data.

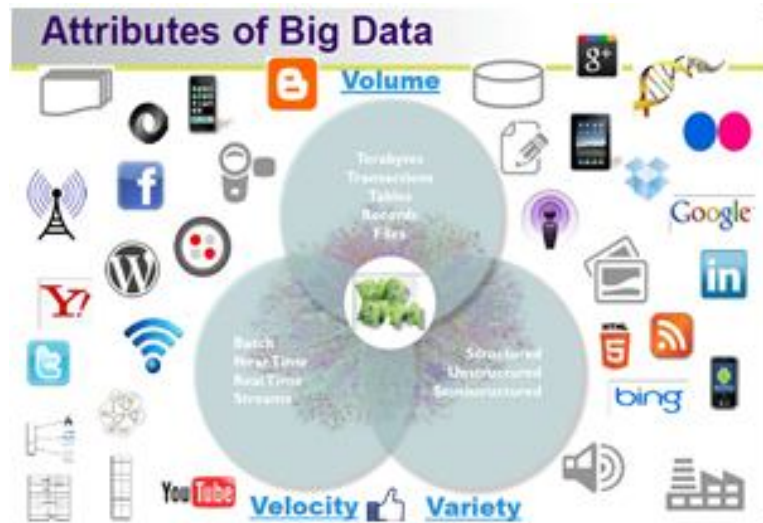


Figure 1: Attributes of Big data

a. Volume

In big data, volume is the most challenging aspect. Big enterprises have large amount of data. The data is in form of system logs, recordkeeping and in other formats as well. Data warehouse based algorithms may not be able to process and to analyze the data effectively due to the lack of an architecture based on parallel processing. We can derive a lot from textual data, location data or log files. Some examples could be email communication patterns, transaction-based data, and security investigation. Big data technology provides a solution which creates a value from this massive data which was difficult to process previously.

b. Velocity

This refers to the speed at which data is flowing in to organization. The speed, at which data is generated and processed, plays a major role in the path of development and growth of big data concepts. In the current era of smartphone, there is even location based data generated and becoming important to be able to process this huge amount of data.

c. Variety

The data which is generated with the social and electronic media is highly unstructured. For example: Unstructured text documents, videos, audios and images. It is required to manage the complexity of data in many different structures, ranging from relational to raw texts.

5. Review of Literature

The privacy issues and the methods to protect privacy for four types of users: data providers, collectors, miners and decision makers [1-2]. Also, some light on the game theory has been shed to understand the model of common grounds for all the users. There are 4 phases of big data – data generation, acquisition, storage and analysis [3-4]. Application of big data is also discussed in this paper which are medical, enterprises, social networks, smart grid etc. [5-6]. The big data and related technologies are - cloud, IOT, Hadoop and data centers [7]. The major challenge of big data and data mining is – Privacy [8-9]. Big data tools are required for successfully managing analyzing and assimilating the structured and unstructured large volume of data produced by the healthcare systems [10]. The main application of the big data is to improve the quality of life of citizens in the cities. So it becomes important to understand the various benefits and challenges that can be faced by big data in smart cities [11]. As the healthcare consists of electronic devices as the source of big data, security and privacy becomes a major concern. There are several international organizations, such as HIPAA, which design regulations for big data [12]. Though the data analytics is growing largely in healthcare sector, it still lacks in some contexts, such as guarantee for the privacy, safeguarding security and establishing standards and governance. Big data analysis and applications in healthcare are at an early phase of development, yet quick advances in stages and apparatuses can quicken their developing procedure [13]. Some ML (Machine Learning) concepts are also being applied for big data analysis in the process of taking effective decisions in healthcare domain [14-16]. In order to help clinical decisions, a big data based knowledge management system has been proposed which is based upon variety of databases such as EHR (Electronic Health Record), Medical Imaging Data, Unstructured Clinical Notes and Genetic Data etc. [17].

6. Conclusion

This survey aims at studying and understanding of the data analysis of skin diseases using big data analytics. The study is about how to analyze the datasets of different skin diseases and techniques to provide the symptoms and prescription of the particular disease. Data analysis of skin diseases using big data analytics is an evolving concept. The process of data analytics are considered as: Data collection, Data cleaning and Data processing. In Big data analytics, 3Vs – Variety, Volume and Velocity- are studied about any data. Some big data tools can also be used to implement and analyze the skin disease datasets.

7. Reference

- [1] BijeshDhyaniAnuragBarthwal. Big Data Analytics using Hadoop, International Journal of Computer Applications (0975 – 8887) Volume 108 – No 12, December 2014
- [2] S Ramyasilpa,chidvila v, A Review of skin cancer. International research journal of pharmacy 2013.
- [3] Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar,A review paper on big data and Hadoop. International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014
- [4] Min Chen, Shiwen Mao, Yunshan Liu, Big Data: A Survey. Springer, January 2014.
- [5] Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, Yong Ren. Information Security in Big Data: Privacy and Data Mining, IEEE Access, October 2014
- [6] Emad A Mohammed, Behrouz H Far, Christopher Naugler. Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends, Biodata Mining 2014.
- [7] Matthew Herland, Taghi M Khoshgoftaar and Randall Wald. A review of data mining using big data in health informatics, Journal of Big Data, 2014.
- [8] Priyanka K, Prof NagarathnaKulennavar, A Survey on Big Data Analytics in Health Care International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014
- [9] Maryam M Najafabadi,FlavioVillanustre,Taghi M Khoshgoftaar,Naeem Seliya,Randallwald and EdinMuharemagic. Deep learning application and challenges in big data analytics, Journal of Big Data, 2015
- [10] Ashwin Belle, RaghuramThiagarajan, S.M.RezaSoroushmehr, FatemehNavidi, Daniel A. Beard and Kayvan Najarian1. Big Data Analytics in Healthcare. Biomed Research International · January 2015
- [11] Eiman Al Nuaimi, Hind Al Neyadi, Nader Mohamed, Jameela Al- Jaroodi, Applications of big data to smart cities. Journal of Internet Services and Applications, 2015 .
- [12] Dr. Smitha Rao, Ms. Suma. S. N, Ms. Sunitha. M, Security Solution for Big Data Analytics in Healthcare. IEEE, 2015
- [13] Dharavath Ramesh PranshuSuraj,andLokendra Saini, Big data Analytics in Healthcare: A Survey Approach. IEEE Access, 2016
- [14] Priyank Jain, ManasiGyanchandani and NilayKhare,Big data privacy: a technological perspective and review,journal of Big Data 2016.

- [15] Allard J. van Altena, Perry D. Moerland, Aeilko H. Zwinderman and Sílvia D. Olabarriaga, Understanding big data themes from scientific biomedical literature through topic modeling, journal of Big Data 2016.
- [16] Jasleen Kaur Bains, Big Data Analytics in Healthcare- Its Benefits, Phases and Challenges, International Journal of Advanced Research in Computer Science and Software Engineering Volume 6, Issue 4, April 2016.
- [17] Gunasekaran Manogaran, Chandu Thota, Daphne Lopez, V. Vijayakumar, Kaja M. Abbas and Revathi Sundarsekar. Big Data Knowledge System in Healthcare. Springer International Publishing AG 2017.