Chapter 51

Big Data Predictive Analysis for Detection of Prostate Cancer on Cloud-Based Platform:
Microsoft Azure

Ritesh Anilkumar Gangwal
Dr. Babasaheb Ambedkar Marathwada University, India

Ratnadeep R. Deshmukh
Dr. Babasaheb Ambedkar Marathwada University, India

M. Emmanuel
Pune Institute of Computer Technology, India

ABSTRACT

Big data as the name would refer to a subsequently large quantity of data which is being processed. With the advent of social media the data presently available is text, images, audio video. In order to process this data belonging to variety of format led to the concept of Big Data processing. To overcome these challenges of data, big data techniques evolved. Various tools are available for the big data naming MAP Reduce, etc. But to get the taste of Cloud based tool we would be working with the Microsoft Azure. Microsoft Azure is an integrated environment for the Big data analytics along with the SaaS Cloud platform. For the purpose of experiment, the Prostate cancer data is used to perform the predictive analysis for the Cancer growth in the gland. An experiment depending on the segmentation results of Prostate MRI scans is used for the predictive analytics using the SVM. Performance analysis with the ROC, Accuracy and Confusion matrix gives the resultant analysis with the visual artifacts. With the trained model, the proposed experiment can statistically predict the cancer growth.

INTRODUCTION

Big data as the name would refer to a subsequently large quantity of data which is being processed. As per Definition: Big Data is a term for data sets processing that are so large or complex that traditional data processing applications are not adequate".

As per (Fouad, K., Hassan, B., & Hassan, M. (2016)) the current scenarios where the capacity of storage devices is increasing constantly from terabytes to petta-bytes to Zettabytes (43 Trillion Gigabytes) so on, it is become difficult for the relational database systems and query processing system to operate. Now the question might arise, with the advent of huge storage space the primary memory is also multiplying thus counterfeiting the problem of huge data, so why is the need for the additional tool for the Big data.

To define the term in some common words is: "Big Data is not only about huge quantity of data, but it is a way of finding new insights into some existing data, to analyze and get the unknown facts". As per (Dave, P.2016 & Fernández et al., 2014), it can make businesses dynamic and robust, so as to overcome various business hurdles and adapt to the growing challenges.

Big data refers to scenarios preciously with the 3V's.

1.   Volume
2.   Velocity
3.   Variety

Volume

With the advent of transistor, there has been an explosive growth in the data storage device's capacity. Thus the data stored is now just more than text. Currently data is available in various forms may it be conventional text, Video streams, Audio's etc. which is most oftenly found in the social media or News Channels. As the storage capacity grows, the applications and architecture already developed also needs to be evaluated again. Sometimes the same data is examined with respect to multiple dimensions, even though the original data is not altered. The big volume thus represents most usual aspect of Big Data.

Velocity

There has been a huge change in perspective of how we look at the data due to huge data growth and social media. Earlier, data available of yesterday was consider recent, but today with the advent of social media, we reply on them in order to provide update with latest happenings. Sometimes, a few moments old data on social media is considered outdated and rendered useless, thus does not interests users. Such old messages are often discarded and people are diverted to the latest or recent update. The data movement has to be real time and update time needs to be reduced to fraction of seconds. Thus the high velocity plays a critical role to represent big data.

Variety

The data is stored in multiple formats in the data storages may be it in the form of database, excel sheets, comma separated files, text file, images, PDF files, video files etc. example, O'Shaughnessy, S. & Gray, G. (2011). This makes it necessary for the organization to manage and make the data meaningful. It is

Related Content

Optimal Fault Tolerance Strategy Selection for Web Services
Zibin Zhengand Michael R. Lyu (2010). *International Journal of Web Services Research (pp. 21-40).*
www.irma-international.org/article/optimal-fault-tolerance-strategy-selection/47041

Data, Information, and Knowledge: Developing an Intangible Assets Strategy
G. Scott Ericksonand Helen N. Rothberg (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 401-412).*
www.irma-international.org/chapter/data-information-and-knowledge/217842

A Specification Framework for Big Data Initiatives
Anh D. Ta, Marcus Tanqueand Montressa Washington (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 639-656).*
www.irma-international.org/chapter/a-specification-framework-for-big-data-initiatives/217854

An Energy-Aware and Under-SLA-Constraints VM Consolidation Strategy Based on the Optimal Matching Method
WeiLing Li, Yongbo Wang, Yuandou Wang, YunNi Xia, Xin Luoand Quanwang Wu (2017). *International Journal of Web Services Research (pp. 75-89).*
www.irma-international.org/article/an-energy-aware-and-under-sla-constraints-vm-consolidation-strategy-based-on-the-optimal-matching-method/188458

Selective Service Provenance in the VRESCo Runtime
Anton Michlmayr, Florian Rosenberg, Philipp Leitnerand Schahram Dustdar (2010). *International Journal of Web Services Research (pp. 65-86).*
www.irma-international.org/article/selective-service-provenance-vresco-runtime/42110