Chapter 19
Future Expectations About Big Data Analytics

Santosh Ramkrishna Durugkar

https://orcid.org/0000-0002-5079-2224

Independent Researcher, India

ABSTRACT

There is no future without big data as users, applications are continuously increasing and processed information expected to use in Healthcare, e-Commerce, Aviation, Education, etc. One can say it is just the beginning of big-data-era (BDE) computing. Everyone expects that 'ready to use' data must be available instantly and hence new techniques and algorithms must be developed to store handle and find the required 'relevant' data. However, to predict and find the relevant data many researchers suggested taking care about handling the 'context'. Many people process big datasets with missing contexts, and considering the datasets or attributes as per their convenience. In future it is also expected to focus on data quality because many applications will use this ready to process data as an input. Decision making is possible with recent machine learning applications and selection of data. It involves carefully selecting the data and removing incomplete, invalid, inaccurate data. Therefore, it is suggested to get ready for the future of big data with innovative ways, techniques, and algorithms.

1. INTRODUCTION

In many applications big data can be integrated from a variety of sources and termed *heterogeneous data* (Heterogeneous big data). However, in the future, users 'u', and devices 'd' will increase which generates heterogeneous data (text, audio, images, video etc.) that requires handling & retrieving important information. *Future expectation from big-data processing applications requires developing applications based on speedy algorithms representing this heterogeneous data in aggregate form*. However, a centralized coordinator will be necessary if the system involves multiple users, multiple systems etc. In the future, *proper coordination* is the key challenge in handling and processing the big-data. In the future focus should be on developing applications to handle '***contextual data***'. These applications can be termed as *Next-Generation-big-data* (NGBD) processing contextual data. Applications of context

data are traffic handling, weather forecasting etc. Therefore processing historical as well as current data requires developing NGBD applications. *We strongly believe these types of systems will serve the next generation of businesses.*

If observing the *trend of digitization*, one can conclude that 'n' users are increasing and the simultaneous increase in the 'n' users added large volume of data in the system. However, this large volume of data is heterogeneous because it is integrated from different sources. Transformation of the data in unique representation is necessary for analysis purposes. Big data analytics helps users fulfilling their daily business needs, identify the customer segment and forecast business strategies. Traditional learning does not support in-depth learning whereas recent machine learning (ML), deep learning (DL) uses supervised & unsupervised learning techniques. Though the datasets are large & consist of 'n' records/values using classification methods one can classify the features of the datasets. Among those records/values, one can select a few records as a representative value instead of considering all the records from the given dataset which is known as *sampling*.

To store this large volume of data on a bigger platform 'cloud' service is necessary and one can go for *open source cloud service providers*. Though, this big data resides on the cloud there could be some risks associated with the storage. These risks can be mitigated using the various applications such as *Cassandra service* if there is any issue with the data and safe location can be determined by querying the measurable property such as write/read errors with the disks. A future expectation about big data gives more suggestions like:

- *User must use the data more efficiently & effectively*
- *To process the large volume of data more sophisticated algorithms are required*
- *Along with these parameters there is a more requirement of more intelligent and fast machines.*

In future user's expectations will definitely increase i.e. Algorithm, Machines must process large volume of data and works on '*personalization*'. According to individual users will expect categorizing the big data into different categorizes i.e. user's specific data. This approach will give rise to *predictive analytics*. Another future expectation is migrating the large volume on the cloud. Due to rapid growth of users and data driven applications data size will also increase. Even though users will store the large volume on the cloud, it must incorporate *'data security', 'data integrity' and 'data availability'* i.e. SIA model must be incorporated in big data driven applications. Users will expect processing the large volume of data on the cloud & make it available as and when required. In this case HADOOP, NO-SQL etc. platforms can be very useful. *Transformation of large volume of data on the cloud is possible with the use of AWS, Microsoft Azure, Google Cloud platforms.* As discussed earlier, with the increasing users, application's data size will also increase and requires *skilled persons* to process this big data. These skilled persons will store, process, and analyze the large volume of data. Therefore in future *data analysts, data administrators, data scientist* roles will be high in demand. Many data driven applications and systems works on the '*data*' and it is expected future is 'data driven' which solves many business problems and will definitely help to increase the growth of businesses.

Big data could be the next basic need to enhance the performance and growth of the businesses, retaining the customers and personalizing the user's choices in ecommerce industries etc. Businesses will keep track of their customers using this 'raw data' (big data) to classify their choices, feedbacks etc. Few research studies also suggested working on 'real-time' analysis of this large volume can also transform the business, industries, & living of human beings. Even emphasis can be given on 'You-

Related Content

Mathematical Programming Models for Classification

Minghe Sun (2014). *Encyclopedia of Business Analytics and Optimization (pp. 1503-1516).*

www.irma-international.org/chapter/mathematical-programming-models-for-classification/107344

Incorporating the Predictability of Consequences into a Disruption Management Framework

Jason M. Riley, Janis Millerand V. Sridharan (2015). *International Journal of Business Analytics (pp. 20-32).*

www.irma-international.org/article/incorporating-the-predictability-of-consequences-into-a-disruption-management-framework/126831

Analytical Customer Requirement Analysis Based on Data Mining

Jianxin Jiao, Yiyang Zhangand Martin Helander (2006). *Business Applications and Computational Intelligence (pp. 227-247).*

www.irma-international.org/chapter/analytical-customer-requirement-analysis-based/6027

Intelligent IoT-Enabled System in Green Supply Chain using Integrated FCM Method

Rui-Yang Chen (2015). *International Journal of Business Analytics (pp. 47-66).*

www.irma-international.org/article/intelligent-iot-enabled-system-in-green-supply-chain-using-integrated-fcm-method/126833

RezFind: NLP-Based Resume Shortlisting

Srikanth K., Dibya Nandan Mishra, Pratyusha Pujari, Pingili Sravya, Vishal K.and Sankalp Chenna (2023). *Handbook of Research on AI and Knowledge Engineering for Real-Time Business Intelligence (pp. 296-306).*

www.irma-international.org/chapter/rezfind/321501