International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Impact of Big Data Analytics on Business Intelligence-Scope of Predictive Analytics

P. Soumya Sree Laxmi†* and P.Sree Pranathi‡

†Department of Computer Science, St. Pious X Degree & PG College for Women, Nacharam, Hyderabad, India-500 007
‡Sreenidhi Institute of Science and Technology, Ghatakesar, Hyderabad, India-501 301

Accepted 22 March 2015, Available online 29 March 2015, **Vol.5, No.2 (April 2015)**

Abstract

The Big Data is the buzzword in present days across the globe. As there is a huge data generated by different sources such as business people, marketing, education, engineering, medicine, social media, on-line transactions, call centers, sensors, web logs and telecommunication. There is a need to manage the risk associated with the collection, storage, retrieval and analysis of such enormous data. Big data analytics is emerged as a core business practice that firms use to understand their business. Business intelligence and analytics are in high demand as organizations seek to use information assets to improve business outcomes. To meet the demand of the Business intelligence (BI) vendors are looking at predictive analytics to understand customer behavior in new ways. E –Business merchants are collecting and analyzing data to discover shopping patterns that will predict value and understand consumer experiences in digital context. In this paper we described the evolution of different analytics and the impact of big data on business intelligence and also the future scope of predictive analytic with the existing business which will yield a better and cost reduced outcomes.

Keywords: *Big Data, Business Analytics, Business Intelligence, Decision making, E-Business, Predictive Analytics.*

1. Introduction

[1]Organizations of all shapes and sizes today are focused on how they can realize higher value from data. With insights from business intelligence (BI) and analytics, firms can improve customer experiences; make smarter decisions about how to allocate resources, and develop strategies to improve business performance. Data analytics emerging as a core business practice that e-commerce firms, their business and their customers are looking at in a deep new ways. The limitations in analytics are not due to the size of a business. Rather, the main limitation is the quality of the data available. With improved data collection and cleansing methods, and new mathematical methods of analysis of data prediction engines can be more effectively trained and the resulting analytics are more reliable and achievable. We would further look into detail how business analytics was overridden by business Intelligence (BI) and further the evolution of Big data as a solutions for analyzing not just raw structured data, but also semi structured and unstructured data from a wide variety of sources, put together as Big Data Analytics, the hottest emerging practice in BI today.

*Corresponding author **P. Soumya Sree Laxmi** is working as Lecturer and **P.Sree Pranathi** is a B.Tech (ECM) Student

2. Business Analytics

Business analytics (BA) refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning (Bartlett, *et al* 2013). Business analytics is used by companies committed to data-driven decision making. BA is used to gain insights that inform business decisions and can be used to automate and optimize business processes. Data-driven companies treat their data as a corporate asset and leverage it for competitive advantage. Successful business analytics depends on skilled analysts who understand the technologies and the business and an organizational commitment to data-driven decision making. Examples of BA uses include:

- Exploring data to find new patterns and relationships (data mining)
- Explaining why a certain result occurred (statistical analysis, quantitative analysis)
- Forecasting future results (predictive analytics).

3. Business Intelligence

Business intelligence (BI) is an umbrella term that includes the applications, infrastructure and tools, and

best practices that enable access to and analysis of information to improve and optimize decisions and performance says Gartner. It is a type of application software which is designed to generate reports on periodic basis, analyze and represent data. Mostly data is read from data warehouse or data mart. It combines knowledge-driven data mining and method-driven data mining, and fills the gap between business intelligence knowledge and existent various data mining methods in e-Business.

4. Big Data

Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze the information (Tom Whites, 2012).

As the data is increasing voluminously in size from petabytes (1 petabyte=1024 terabytes), exabytes (1 exabyte=1024 petabytes),zettabytes (1 zettabyte=1024 exabytes), yottabytes (1 exabyte=1024 yottabytes).

5. Key Characteristics of Big data

5.1 Volume(Scale): The quantity of data produced today by different sources such as Internet usage, social networks, mobile devices, sensors, embedded systems, and enterprise IT is rapidly growing greater than anything ever seen. Data volume is increasing exponentially, which is a challenge of Big data.

5.2 Variety (Complexity): It is the nature of data that exists within big data. This includes different data formats, data semantics and data structures types.

5.3 Velocity (Speed): It is also about the rate of changes, about linking data sets that are coming with different speeds and about bursts of activities, rather than habitual steady tempo. It is important to realize that events in data arise out of the available data.

5.4 Harness of data: As there is a huge amount of unformatted data generated, one should even know how to harness the data and process it for better decision making. These are the different ways in which the data can be processed.
Operational Database: (OLTP) Online Transaction Processing (DBMSs).
Data Warehousing: (OLAP) Online Analytical Processing.
Stream Computing: Due to uninterrupted networks data is arriving in continuous stream 24/7 a day. This data need to be captured and processed using Real-Time Analytics Processing (RTAP).

Big data isn't just a process for storing petabytes or exabytes of data in a data warehouse, (Wiley, *et al*, 2013). It's about the ability to make better decisions and take meaningful actions at the right time. The biggest challenge for most E-Commerce businesses is to collect, store and organize data from multiple data sources. Big data analytics refers to the strategy of analyzing large volumes of data, or big data. This big data is gathered from a wide variety of sources, including social networks, videos, digital images, sensors, and sales transaction records. The aim in analyzing all this data is to uncover patterns and connections that might otherwise be invisible, and that might provide valuable insights about the users who created it.

6. Big Data Technology

Some the resent technology emerged in Big data are:

6.1 Hadoop: An open source (free) software framework for processing huge datasets on certain kinds of problems on a distributed system. Its development was inspired by Google's MapReduce and Google File System. It was originally developed at Yahoo! and is now managed as a project of the Apache Software Foundation.

6.2 HBase: An open source (free), distributed, non-relational database modeled on Google's Big Table. It was originally developed by Powerset and is now managed as a project of the Apache Software foundation as part of the Hadoop.

6.3 MapReduce: A software framework introduced by Google for processing huge datasets on certain kinds of problems on a distributed system.32 Also implemented in Hadoop.

6.4 Cassandra: An open source (free) database management system designed to handle huge amounts of data on a distributed system. This system was originally developed at Facebook and is now managed as a project of the Apache Software foundation.

6.5 Extract, transform, and load: (ETL)Software tools used to extract data from outside sources, transform them to fit operational needs, and load them into a database or data warehouse.

6.6 Cloud computing: Cloud computing as a computing paradigm in which highly scalable computing resources, often configured as a distributed system, are provided as a service through a network.

6.7 Data warehouse: Specialized database optimized for reporting, often used for storing large amounts of structured data. Data is uploaded using ETL (extract, transform, and load) tools from operational data stores, and reports are often generated using business intelligence tools.

6.8 Data mart: Subset of a data warehouse, used to provide data to users usually through business intelligence tools.

6.9 Google File System: Proprietary distributed file system developed by Google; part of the inspiration for Hadoop.31

7. Big Data analytics

Data analytics addresses information obtained through observation, measurement, or experiments about an occurrence of interest. The aim of data analytics is to extract as much information as possible that is relevant to the subject under consideration. Further (Blackett, *et al*, 2013) classified the entire field of big data analytics into three levels according to the depth of analysis: descriptive analytics, predictive analytics, and prescriptive analytics.

7.1 Descriptive Analytics: It exploits historical data to describe what occurred. For instance, a regression may be used to find simple trends in the datasets, visualization presents data in a meaningful fashion, and data modeling is used to collect, store and cut the data in an efficient way.

7.2 Predictive Analytics: It focuses on predicting future probabilities and trends. For example, predictive modeling uses statistical techniques such as linear and logistic regression to understand trends and predict future outcomes, and data mining extracts patterns to provide insight and forecasts.

7.3 Prescriptive Analytics: It addresses decision making and efficiency. For example, simulation is used to analyze complex systems to gain insight into system behavior and identify issues and optimization techniques are used to find optimal solutions under given constraints.

8. Application Evolution and emerging Analytics

As per the application evolution depicted in the figure bellow, emerging analytics research can be classified into five critical technical areas: data analytics, text analytics, web analytics, network analytics, and mobile analytics. This classification is deliberate to highlight the key data characteristics of each area.

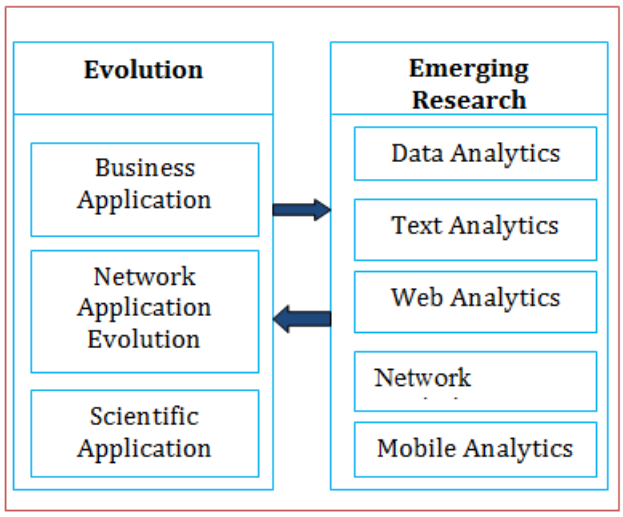

Fig.1 Data Evolution and Emerging Research

8.1 Data Analytics: Business and scientific research fields are generating a large amount of unstructured data. Management of these structured data relies on the proper RDBMS, data warehousing, BPM, and OLAP. Data analytics is largely grounded in data mining and statistical analysis. A new method, process mining has become an intensive transformation and analytics. Hadoop- and MapReduce-based systems have become another viable option for big data analytics in addition to the commercial systems developed for RDBMS, column-based DBMS, in-memory DBMS, and parallel DBMS (Chaudhuri *et al*, 2011).

8.2 Text Analytics: The significant portion of unstructured content collected by firms is e-mail communication, corporate documents, web pages, and social media content. Hence, text analytics is believed to have higher commercial potential than structured data mining. In general, text analytics, also known as text mining, refers to the process of extracting useful information and knowledge from unstructured text. Several technologies have been developed for text mining, including information extraction, topic modeling, summarization, categorization, clustering, question answering, and opinion mining. Opinion mining is a computational technique for extracting, understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated contents.

Similar to big data analytics, text analytics uses MapReduce, Hadoop, and cloud services to continue to foster active research directions in both academia and industry.

8.3 Web Analytics: Web analytics with the popularity of Web 2.0 systems, has grown abundantly. Web analytics aims to retrieve, extract, and evaluate information for knowledge discovery from web documents and services automatically. The content is both pushed in and pulled out in various customer related transactions. HTTP/HTML-based hyperlinked web sites and associated web search engines and directory systems for locating web content have helped develop unique Internet- based technologies for web site crawling the web page, updating, web site ranking, and search log analysis. The increase multimedia content also is a reason of research in web analytics.

A major emerging component in web analytics research is the development of cloud computing platforms and services, which include applications, system software, and hardware delivered as services over the Internet. Based on service- oriented architecture (SOA), server virtualization, and utility computing, cloud computing can be offered as software as a Business Intelligence Research service (SaaS), infrastructure as a service (IaaS), or platform as a service (PaaS). Only a few leading IT vendors are currently positioned to support high-end, high-throughput BI&A applications using cloud computing.

8.4 Network Analytics: As there is a rapid growth of online social networks, network analysis has evolved from earlier bibliometric analysis and sociology network analysis to the emerging social network analysis. Typically, social networks contain a tremendous amount of linkage and content data, where linkage data are essentially the graph structure, representing communications between entities and the content data contains text, images, and other multimedia data in the networks. However many techniques have been developed to study the dynamic nature of social networking.

From the data-centric view, there are two primary research directions in the context of social networks: Linkage-based structural analysis and content-based analysis. Linkage-based structural analysis focuses on areas of link prediction, community detection, social network evolution so on. The term social media is employed to name such user generated content, including blogs, photo and video sharing, social book marketing, social networking sites, social news and wikis. Social media content contains text, multimedia, locations and comments. In today E-Content generation almost every research topic on structured data analytics, text analytics, and multimedia analytics can be translated to social media analytics.

Social media analytics face some current challenges. One ever-growing data in social media, two data produced may be noisy data and lastly data is ever changing, updating with respect to rising technology. Clustering is accomplished by determining sets of nodes with similar content, as social networks contain a large amount of linked information among different types of objects.

8.5 Mobile Analytics: Mobiles became an effective channel for reaching many users and as a means of increasing the productivity and efficiency of an organization's workforce. Mobile BI was also considered by the Gartner BI Hype Cycle analysis as one of the new technologies that have the potential to drastically disrupt the BI market (Bitterer, 2011). With the rapid growth of mobile computing and more mobile terminals like mobile phones, sensors, RFID and applications are deployed globally. Recent advances in wireless sensors, mobile technologies, and streaming processing have led to the deployment of body sensor networks for real-time monitoring of an individual's health.

The lightweight programming models of the current web services (e.g., HTML, XML, CSS, Flash, Ajax) and the budding mobile development platforms such as Android and iOS have contributed to the rapid development of mobile web services.

The trend that is important to understand in the context of cloud computing and authentication is the shift in platforms from traditional PCs toward smart phones. In the IT, (BYOD) Bring your own device, is a phrase that has become widely adopted to refer to employees who bring their own computing devices – such as smart phones, laptops and PDAs – to the workplace for use and connectivity on the secure corporate network.

In addition to all these the trend of online, web marketing through mobiles has contributed to the development of e-commerce apps and social apps according to IBM survey.

9. Future of Predictive Analytics

Using predictive analytics, organizations have a new way to obtain real-time, data-driven insights about what the future may hold. They can leverage sophisticated statistical analysis techniques to mine their data, find what factors can impact their business, and build models that will simulate what will happen when certain conditions arise. These models can be used to conduct what if analyses and provide new insight so organizations can proactively manage their business objectives.

Predictive analytics identifies the meaningful patterns of Big data to predict future events and access to various options. Predictive analytics can be applied to any type of unknown data, whether it is past, present or future related data. Predictive analytics provides the Business Intelligence about the future using insight of big data.

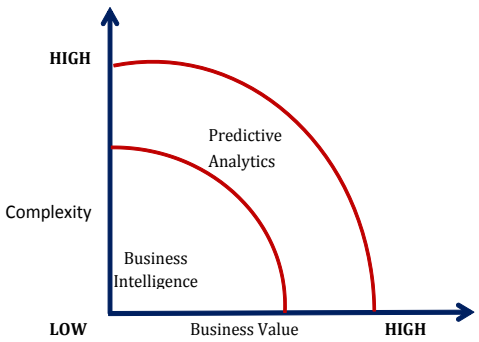

Fig.2 Difference between Business Intelligence and Predictive Analytics

10. Benefits of big data in predictive analytics

- A wide variety of data from different sources can be captured, stored, and processed. More real-time analytics may be conducted from desktops or mobile devices for ad hock decision making.
- Large volumes of non-transactional data can be included in analytics.
- The time to solution is significantly shortened.
- Future possibility of automation of data modeling.

Conclusions

The field of predictive analytics is the next evolution in business intelligence - moving beyond the practice of creating reports on past events, and towards the use of sophisticated statistical methods to predict future outcomes.

Three things are required to implement predictive analytics:

1) Access to the right historical or real-time data that can be used to uncover predictive factors.
2) A solid understanding of the statistical techniques necessary to investigate the data and develop predictive models.
3) Access to the necessary mathematic and statistical algorithms and infrastructure.

Further many a new trends in predictive Analytics is making a way in to E-business to understand customers behavior as a concept of personalization.

References

Tom Whites, (10 May 2012), Hadoop: The Definitive Guide. O'Reilly Media. p. 3. ISBN 978- 1-4493-3877-0.

Michael Minelli, (Wiley, 7 Mar 2013), Big Data Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses.

Bartlett, Randy (February 2013), A Practitioner's Guide to Business Analytics: Using Data Analysis Tools to Improve Your Organization's Decision Making and Strategy. McGraw-Hill. ISBN 978-0071807593.

G. Blackett. (2013), Analytics Network-O.R. Analytics [Online].http://www.theorsociety.com/Pages/SpecialInterest /AnalyticsNetwork_anal%ytics.aspx.

Chaudhuri, S., Dayal, U., and Narasayya, V. (2011), An Overview of Business Intelligence Technology, Communications of the ACM (54:8), pp. 88-98.

Bitterer (2011), Hype Cycle for Business Intelligence, Gartner, Inc., Stamford, CT.