Predictive Analytics in Data Science for Business Intelligence Solutions

Parth Wazurkar
Dept of CSE,
Indian Institute of Information Technology, Nagpur,
Maharashtra, India
parthwazurkar@gmail.com

Robin Singh Bhadoria
Dept of CSE,
IIIT, Nagpur
Maharashtra,India
robin19@ieee.org

Dhananjai Bajpai
KFX Circuits and Systems,
Bengaluru, India
dhananjai.b@kfxlabs.com

***Abstract* –In modern era of computing, organizations are focusing on the better utilization of technology and surviving to gear-up with global business demand. Such competition is acting as a driving force for its business to cope-up the data which generated every second of minute. This data needs to figure out and segregated with information which is required for is business growth model. The Predictive Analytics (PA) uses various algorithms to find out different patterns in large data that might suggestthe efficient behavior for business solution. This paper provides a conceptual decision making process for data using predictive analysis to maximize the success ratio for handling large dataset. Today, different technologies like cloud computing, SOA, are together transforming information technology but in turn, are imposing new complexities to the data computation. Due to such advances in technologies, and itrequires rapid and dynamic data analysis for structured and unstructured data.**

Keywords – Predictive Analysis,Large Dataset, Business Intelligence Solution, Data Analysis.

1. I. INTRODUCTION

After the principles of Agile Software Development (ASD) were published there has been a change in Business Intelligence as the objectives and principles of Agile Software Development have been applied to Business Intelligence (BI) which has led to a lot of development in this sector. BI is referred as the techniques or practices which utilize different technologies to create different methods or applications which analyze the business data available with the organization to help the enterprise to take decisions based on the predictions made by the data. BI not only includes the data processing and analytical technologies but also many business centric practices and methods which can be applied to various applications such as e-governance, health-care, e-commerce, security and market intelligence [13]. There has been a lot of development in BI, which has led to a lot of applications.

This article provides the way of applying the agile principles to BI delivery, fast analytics, and data science. The core ideals: individuals and interactions over processes and tools; working software over comprehensive documentation; customer collaboration over contract negotiation; and responding to change over following a plan of the manifesto. These ideals have made the software development less formal, more dynamic, and customer focused. Information Technology (IT) departments are facing a challenge of maintaining a competitive edge, which, in turn, is increasing pressure for delivering high quality technology solutions faster. Under these circumstances, the accurate measure for measuring the value of technology efforts is how soon payback and return on investments occur.The measurement of BI value continues to be a struggle for many organizations, mainly due to the challenge of attributing return to the investment and overall performance. The BI enables the organization to become smarter, work smarter, and helps it to take better decisions through the use of information.After the information has been extracted from the data the information is yet to be interpreted the process used to interpret and derive value from information is often called as information value chain. The first step in the value chain is the extraction of data from different sources; applying different logics and business contexts to this data creates information; information is then consumed by BI users; Based on these information different decisions are made and executed; thus increasing the business value.

Business Intelligence (BI) has been defined by literature and scholars in much similar ways. BI has been able to improve the success of organization by providing better decision making with the use of information which the regular reporting did not provide.BI requires different tools, applications, and technologies focused on enhanced decision-making which is commonly used in supply chain, sales, finance, and marketing [12]. BI is a process which applies proper knowledge or intelligence to data to extract information and applies it into business decision making process. BI helps organizations to improve the decision making process and which otherwise requires different processes, skills, technology, and data. (One of the major challenges faced by BI is the better collaboration between business and Information Technology which actually results in creation of information the raw data. Some of the hurdles faced in any BI project includes: the lacking in understanding about how data is created and used to turn into information, the data available with us is not quantifiable based on quality; results are not demonstrated in a timely manner; and the lack of trust between IT and

978-1-5386-1860-8/17/$31.00 ©2017 IEEE
DOI 10.1109/CSNT.2017.70

Authorized licensed use limited to: INDIAN INSTITUTE OF INFORMATION TECHNOLOGY. Downloaded on May 05,2023 at 21:08:27 UTC from IEEE Xplore. Restrictions apply.

business stakeholders . While combating these challenges, the need to obtain information sooner has been influenced by the phenomenon of "Big Data" [2, 6].

2. II. Background Works

As BI is a data-centric approach, it depends heavily on the database management field.Thus the regular improvement in techniques for data collection, extraction and its analysis has created a direct impact on BI. (Companies collect a lot of structured as well as unstructured data on a regular basis which they store into relational database management systems (RDBMS) [4, 8]. The analytical methods which are commonly used in these systems, were popularized in the late 1990s, and are mainly based on statistical methods and various data mining techniques.

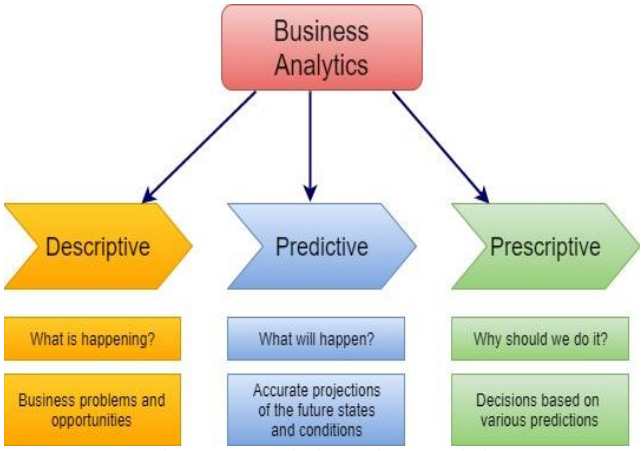

Figure 1. Scenario for Business Analytics

Business intelligence is considered as set of concepts and methods to improve business decision making by using fact-based support systems. The first productive BI systems were implemented at large consumer goods manufacturers for the purpose of analyzing sales data.These traditional BI solutions were mainly focused on analyzing historical data, like for determining the amount of yield of a particular product in certain region and the profit made during a fixed period of time. In the early 2000s, the term "big data" started making a place into scientific literature and today it has become a common word of speech which is used by people on a regular basis, back then "big data" usually referred to data which was too large to be accommodated into local disks or even hard drives [7].

The first publication about big data was originated from the field of scientific computing, which were later conceptualized for business development model. After the mid-2000s, businesses started growing interest in big-data, the started analyzing this data using a variety of algorithms and methods. Companies started exploiting this "big-data" to analyze different problems and developing a solution to solve these problems by applying sophisticated machine learning and data mining techniques.

Today's big data analytics when compared to traditional business intelligence applications, it not only goes deeper into the breadths and depths of data but also tries to answer various different questions which the. While BI traditionally focused on using a predetermined set of methods to measure the past business performance, big data applications emphasizes more on exploration and prediction of different results.

3. III. Predictive Analysis In Business Solution

A large shift towards Big Data from traditional approaches has been observed to handle different business processes and to develop better predictive models for the organization. Business intelligence and analytics is helping many companies to improve their efficiency in customer satisfaction.Predictive modeling has been one of the major reasons in drastically changing the products and services provided by companies in recent years [5].Google search has drastically overtaken most of its competitors and this has been possible because of Google's investment into predictive analysis as it uses different algorithms and various predictive models to predict users' search results and news feeds to better facilitate the user. Amazon also relies on predictive models of what kind of product an user might purchase and how can they manipulate the user to buy the product. The advertisements which are often displayed on user's screen while visiting a website is mainly based on different predictive model which helps the company to better popularize amongst people who could be potential buyers.The applications of predictive algorithms are not only limited to the online world. Health care industries are also transiting towards better utilizing it to provide quality services to humanity [6]. The predictive models based on the data of individual health costs and outcome provides a "risk score" which improves costs and quality of health care.

Predictive analytics tries to predict behaviour in future by finding patterns in the data available with it by applying various different algorithms.If the model results are found to successfully to predict it, then the company will try to find out another solution so to make customer not to churn from their network thus, predictive analytics is a continuous process [1].

To maximize the success of the organization with predictive analytics, following steps must be followed by an organization:

Identifying Business Goals: First step is to identify the business goal that clearly defined successful predictive patterns. For example-business goal might be to improve the suggestion system of suggesting different items to the customer dynamically when the customer is adding products to his shopping cart. It will thus help in increasing the sales of the product, improving the profit of the organization and

Authorized licensed use limited to: INDIAN INSTITUTE OF INFORMATION TECHNOLOGY. Downloaded on May 05,2023 at 21:08:27 UTC from IEEE Xplore. Restrictions apply.

in turn reducing the efforts of the customer has to apply in finding similar items or what he wants to purchase more. This will lead to improvement in customer satisfaction [2].

Data Understanding from Various Sources: After setting up the business goal, the further step is to collect data from variety of sources available.Data can also be collected from external sources for analysis purpose. These external sources can be government data, social media websites, public sector data and many other sources. This data collected from variety of sources helps in augmenting internal data.Data visualization tools can help data analysts to explore data from variety of sources to determine which data is relevant for predictive purpose [9].

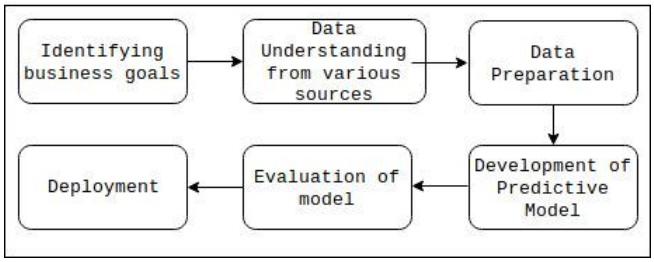

Figure 2. Scenario for Predictive Analysis in Data Science

Data Preparation: The main challenge faced is to prepare the data for predictive analysis as raw data cannot be directly utilized for analysis. Preprocessing must be performed by the analyst so as to get the data ready for predictive analysis [13].

Development of Predictive Model: Data analysts use one or more of the predictive analytics modeling tools to perform various analysis. Various machine learning algorithms and a lot of statistical algorithms are applied by data analysts to devise better predictive models [10].

Evaluation of Model: Predictive analytics is all about probabilistic resulting and not absolutes. A probabilistic model is set up so as to be compared with the outcome of predictive analysis so as to quantify the outcome of the analysis and evaluate its efficiency better. If the predictive output is found to be more effective than the randomly selected output, and then this model is effectively termed as a better predictive model. Data analysts can run various different algorithms to find the most predictive model amongst the different models.If no results are found then it is assumed that data is not suitable for prediction or is not enough to perform predictive analytics [12].

Deployment: Once an effective predictive model is identified, then the only step remains is the deployment of this model in the production application by the analysts. This model consists of methods to run predictive rules for acquiring data as well as obtaining results from the model.

Effectiveness of Model and Result Analysis: It is very much important to continuously evaluate the effectiveness of the model as it might happen that the organization had performed predictive analytics on previously selected data sets but as the market scenario had changed over time the results obtained now might not be favorable. Organizations must continue to perform predictive analytics process so as to firmly stand in the competitive market.

4. IV. Result Analyses for Effective Business Solution

A lifecycle is the development growth of particular and provide detail analysis for specific event monitoring. Such event may report with end of life or finished due to no existence in value. This growth could be monitored with respect to patterns that support that particular event. This could help in monitoring the BI eventsassociated with data analysis. Table 1 specifies the comparison for events based on BI and Predictive analytics [3, 5].

TABLE I. COMPARISON FOR BUSINESS SOLUTION OVER PREDICTIVE ANALYSIS ON DATA SCIENCE PROJECTS

Intelligent Business Life Cycle	Predictive Analytics Life Cycle
Finding	Choice
Plan & Policy	Pattern Analysis
Progress	Heterogeneity
Check	Risk
Deploy	Endorse
Support	Provision

5. V. Conclusion

This paper presents the important facts about predictive analysis and its associated parameters that supports in analysis large dataset. The solutions based on BI are only provides clustering for large dataset but PA offers detail evaluation. Such PA segregates the data based on its generation pattern, development & deployment model that is most effective on statistic datasets.

6. Reference

[1] D. Larson, and V. Chang, "A review and future direction of agile, business intelligence, analytics and data science", International Journal of Information Management, Vol. 36, No. 5, pp.700-710, 2016.

[2] K.S. Jadon, R.S. Bhadoria and G.S. Tomar GS, "A Review on Costing Issues in Big Data Analytics", In Proc. of IEEE International Conference on Computational Intelligence and Communication Networks (CICN), pp. 727-730,Dec 2015.

[3] O. Müller, I. Junglas, J. vom Brocke, and S. Debortoli, "Utilizing big data analytics for information systems research: challenges, promises and guidelines", European Journal of Information Systems, Vol. 25, No. 4, pp.289-302, Jul 2016.

[4] S. Mazumder, R.S. Bhadoria, and G.C. Deka, Distributed Computing in Big Data Analytics, AG: Springer International Publishing, 2017.

[5] A. Gandomi andM. Haider,"Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management, Vol. 35, No. 2, pp. 137-44, April 2015.

Authorized licensed use limited to: INDIAN INSTITUTE OF INFORMATION TECHNOLOGY. Downloaded on May 05,2023 at 21:08:27 UTC from IEEE Xplore. Restrictions apply.

[6] G.S. Tomar, N.S. Chaudhari, R.S. Bhadoria, and G.C. Deka, The Human Element of Big Data: Issues, Analytics, and Performance, FL: CRC Press, Sept. 2016.

[7] M. Bichler, A. Heinzl and W. M. P. van der Aalst, "Business Analytics and Data Science: Once Again", Business & Information Systems Engineering, Vol. 59, No. 2, pp. 77-79, April 2017.

[8] A. Abbasi, S. Sarker, R.H. Chiang,"Big Data Research in Information Systems: Toward an Inclusive Research Agenda", Journal of the Association for Information Systems, Vol. 17, No. 2, Feb 2016.

[9] V. Dhar, "Data science and prediction". Communications of the ACM, Vol. 56, No. 12, pp.64-73, Dec 2013.

[10] D. E. Brown, A. Abbasi, & R. Y. K. Lau, "Predictive analytics: Predictive modeling at the micro level", IEEE Intelligent Systems, Vol. 30, No. 3, pp. 6-8, May 2015.

[11] M. Swarnkar and R.S. Bhadoria, Analysis for Security Attacks in Cyber-Physical Systems. In Cyber-Physical Systems: A Computational Perspective, Eds. G.M. Siddesh, G.C. Deka, K. G., Srinivasa, and L. M. Patnaik, FL: CRC Press, Oct 2015, pp. 489–514.

[12] M.A. Waller and S. E. Fawcett, "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management", Journal of Business Logistics, Vol. 34. No. 2, pp. 77-84, June 2013.

[13] G. Shmueli, and O.R. Koppius,. Predictive analytics in information systems research. MIS Quarterly, Vol. 35 No. 3 pp. 553-572, Sept. 2011.

[14] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin, "Bayesian data analysis", Vol. 2. Boca Raton, FL: CRC press Sept. 2014.

Authorized licensed use limited to: INDIAN INSTITUTE OF INFORMATION TECHNOLOGY. Downloaded on May 05,2023 at 21:08:27 UTC from IEEE Xplore. Restrictions apply.