Predictive Analytics: Benefits, Applications and Challenges

Ms. Uma Pratap Madje
Department of Computer Science
Assistant Professor, Fergusson College (Autonomous), Pune
Uma.madje@fergusson.edu

Abstract — The massive increase of data over the last several years has fuelled the analytics movement. To achieve better informed business decisions with the help of the historical or past data exploration, analytics helps in gaining the business insights. The paper describes the purpose of predictive analytics use in many area of industries to solve the problems and gain the benefits in terms of reduction and prevention of risks, saves time, cost and resource management. This paper present different predictive approaches adapted for different application with challenges and suggestions.

Keywords — Big data, Predictive analytics, Predictive approaches, Challenges

1. INTRODUCTION

Predictive analytics comes from an umbrella that covers the system known as Business Intelligence. BI deals with enhancements of decision making with the help of transformed data and knowledge extracted by using data mining tools, techniques. Predictive analytics mainly deals with analysing the data and predicting the outcomes what the future holds. In other words, it deals with the prediction of future events based on the historical data, observations made on that data by applying some techniques and algorithms. So, one can formally define the predictive analytics as, the area of data mining that deals with exploring, extracting information based on historical data and applying algorithms and techniques to predict the trends and outcomes The prediction process can be generalized in some basic steps:

- Collection of initial or raw data
- Preparation of data, including cleaning, selection, constructing and formatting
- Selecting Modeling techniques i.e. using transformed data
- Report predictions to user using Modeling techniques.

[1] Paper describes the data mining as process to collect, filter, prepare, analyse and store data that will be used to create useful knowledge and supporting the data analytics and modeling techniques. The main concept of analytics is divided into following types:

- Descriptive Analytics: It describes the situations and answers the queries about what is happening?
- Predictive Analytics: This describes the situation about future events and answers about what will happen in future?
- Prescriptive Analytics: It prescribes about what is right choice or solution for given defined problem set.

The process of predictive analytics is encompasses of different phases mainly defined as, problem identification, collection and data preparation, data visualization techniques and model development process. [2]

Predictive analytics can be defined as prediction of the future (based on the historical data to perform analysis) to analyse the historical data and discover the patterns, relationships among these data. While [3] add that the predictive analytics is helpful in organizational structures to predict risks, tendency, and financial increments in terms of revenue enhancements by applying some strategic corrections with the help of structured and unstructured data.

2. MODELS AND ALGORITHMS

Predictive analytics uses different algorithms to generate models. Depending on the pattern discovery, dependent variables, explanatory variables and resultant outcome predictions, various algorithms are used. Few are listed below:

2.1 Classification:

It is one of the data mining technique that is used to evaluate categorical known data. In order to form predictions about for the class labels of unknown data. It classifies the data for model construction based on the training and testing sets, and uses the class labels to classify attributes.

2.2 Clustering:

Clusters are made up of known data and class labels. Each cluster contains similar observations and data. Mostly this technique is beneficial when unknown relationships needs to be discovered.

2.3 Association Rule:

It identifies the events that occur together or in a sequence i.e. helps to find the association between observations. An example of association rule is market – basket analysis which is one of the modeling technique that describes if a customer buys a specific product or set of items he will more or less likely to buy another product or set of items associated with that product.

2.4 Regression:

It is a statistical analysis technique that can be used to model the relationship between continuous variables i.e. the analysis provides the relationship among one or more independent variables or predictors. Usually the outcome of this technique is numeric that is target (prediction) based on one or more predictors.

Based on different process models used for data processing and acquiring knowledge from this data, the most widespread models is the Cross-Industry Standard Process for Data Mining (CRISP-DM) [4]. Figure 1 shows the CRISP-DM according to Reinhart et al. [5].

International Journal of Research in Advent Technology, Special Issue, ICATESM 2019
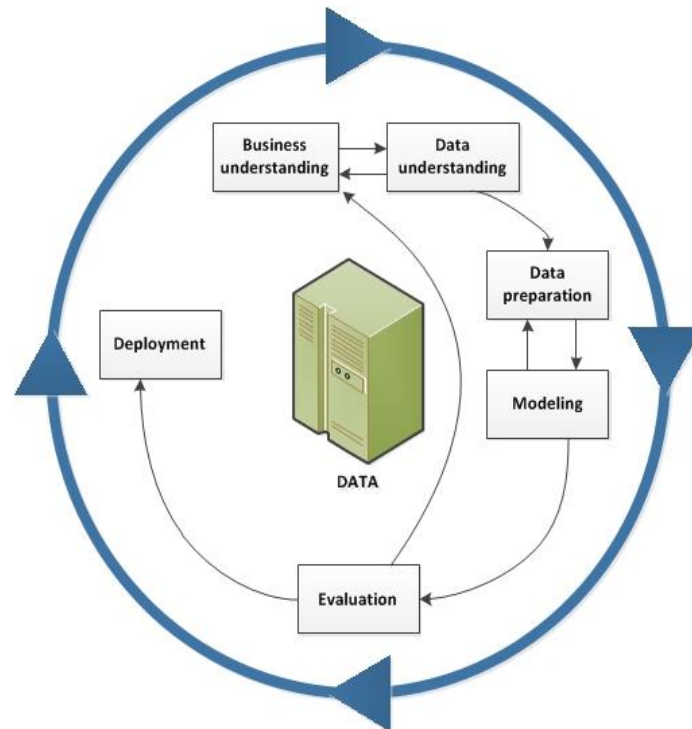*E-ISSN: 2321-9637*
Available online at www.ijrat.org

Fig. 1 CRISP – DM life cycle

CRISP-DM, which stands for CRoss-Industry Standard Process for Data Mining, is an industry-proven way to guide your data mining efforts. As a methodology, it includes descriptions of the typical phases of a project, the tasks involved with each phase, and an explanation of the relationships between these tasks. This iterative process begins with understanding of data and then preparing it for further processing. Following the data pre-processing step, CRISP DM evaluates the data and results in better data understanding. This iterative process leads to model development and deployment in application context.[9]

3. APPLICATION OF PREDICTIVE ANALYTICS IN DIFFERENT INDUSTRIES

The general idea behind Predictive analytics is to detect the relationships and patterns in data by analysing the history of it and based on that information it helps to predict the future for taking preventive decisions. Thus, the main aim of predictive analytics differs from according to the domain / areas in which it's been applied. Ranging from a customer performing an advertisement campaign or product sell to a bank manager who wish to recognize a fraudulent customer or most profitable customer. Hence predictive analytics can be used in many applications. To cite few examples where it made a positive impact:

3.1 Prediction in Higher Education:
In today's world, higher education is been operated at very complex and competitive environment, which leads to facing more challenges accordingly. Not only has that, more stake holders in education paved for big data to play an important role. [4] Suggest that vast data that keeps coming every day, can be utilized only with big data and therefore to capture, store, distribute, manage and analyse big data we require accurate analytic techniques. The effect of big data in higher education depends on making decision about theory from administrative as well as operational perspective and can be processed and accessed to predict future performance and identify potential areas in academic programming, research teaching and learning. Big data can have important role in three data models: descriptive analytics analyse the raw data received and predictive analytics tries to figure out future probabilities based on predictive analysis and prescriptive information are given to students and stake holders [7]. In same context, online learning research community should also bring transparency to effective practices of learning analytics to detect use of big data in it.

3.2 Prediction in Health Care:
The major challenge in health care is quality of service and cost at which the service is rendered. Important constraint is, quality treatment should not backfire on the cost of treatment. Big Data Analytics, which has the potential to improve health care, save lives and lower cost. Understanding patterns, trends within the big data helps to detect the disease at early stage and perform the treatment on it. Based on these detected patterns and in combination with structured and un-structured data, effective care should be given to an individual. In healthcare use of predictive analytics helps to reduce identifying high risk patients and monitoring them at early stage. [8]

3.3 Prediction in Data Governance:
Data governance basically deals with usability, availability and applying security principles to data. [9] Suggest that data governance prevent unauthorized data access, preserving the quality of data by providing the privacy. To increase confidence level, speed with which data needs to be accessed

and performing the faster decision making with precise information in big data, data governance is essential. Ranging from data extraction to secure delivery of data including analysis, maintenance and processing / computing, all these are considered to be as different steps which can be performed in data governance. Due to change in market and business capabilities, information technology has changed a way to treat the data and it functionality or processing. With the integration of available heterogeneous data, technology has made a impact on how a data can be accessed with the inclusion of all security and pricy constraints required for the same.

3.4 Prediction in Customer Orientation:

To place customer satisfaction at core of each business process and taking decisions based on customers benefits, such organizations are termed / coined as Customer-Oriented organization. The customer orientation is an approach to sales and customer relations in which employees are focused on helping their customers to meet their long term needs and wants. Collection of data related to such customers provides behavioral insights and marketers utilize this data, identifies patterns in them to increase the productivity of their business. With the help of information technology, generated patterns and data associated with it can be used for both transactional as well as behavioral purpose. [10] observed and suggest that, large amount of population are spending most of its economy on shopping and purchases, so considering that sort of data, those affluent customers who purchase more expensive products frequently. These collected datasets are temporal information, which can be further observed and used by the sellers for communication among self-regarding the purchase of particular product / item, which helps them in predicting the future purchase of relative customer and making decisions for sale increase / decrease.

4. CONCLUSION

Performing analytics have become a commonplace for wide range of users. The Paper highlights the main application areas depending on big data and incorporating the solution with predictive analytics. The use of business intelligence for modelling and forecasting can be useful in social media analytics. Instead of focusing on specific use case and computing resources, the use of existing analytic framework can be enhanced for huge data sets in diverse areas. Therefore the future implications will be more focused on pattern identification and use of evolutionary techniques from / on various data sources.

REFERENCES

[1] Mohammad Ahmad Alkhatib. "Analysis of Research in Healthcare Data Analytics". 2015. Australasian Conference on Information Systems.

[2] HodaMoghimi, Stephen Vaughan, Steven McConche, NilminiWickramasinghe." How Do Business Analytics and Business Intelligence Contribute to Improving Care Efficiency?" 2016. 49th Hawaii International Conference on System Sciences

[3] PrasadaBabu, S.HanumanthSastry." Big Data and Predictive Analytics in ERP Systems for Automating Decision Making Process". 2014. IEEE

[4] Shearer, C., "The CRISP-DM Model: The new blueprint for data mining", Journal of Data Warehousing, volume 5, pp. 13-22, 2000.

[5] Reinhart, F.: "Industrial Data Science – Data Science in der industriellen Anwendung. Industrie 4.0 Management, 32:27-30, 2016

[6] Ben K. Daniel, "Big Data and analytics in higher education: Opportunities and challenges", British journal of educational technology. September, 2015.

[7] S. Banumathi, Aloysius Assistant Professor, Department of Computer Science, St. Joseph's College, Trichy-2. Tamilnadu, India, "Predictive Analytics Concepts In Big Data- A Survey"

[8] YiChuan Wang, LeeAnn Kung, Chaochi Ting, "Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care", publications on ResearchGate , 2015.

[9] https://www.ibm.com/support/knowledgecenter/SS3R A7_15.0.0/com.ibm.spss.crispdm.help/crisp_overvie w.htm

[10] C. Mohanapriya, "A Trusted Data Governance Model For Big Data Analytics", Volume 1, Issue 7, ISSN (online): 2349-6010, Dec 2014.

[11] Farshad Kooti,Kristina Lerman, Luca Maria Aiello, Mihajlo Grbovic, Nemanja Djuric,Vladan Radosavljevic," Portrait of an Online Shopper: Understanding and Predicting Consumer Behavior" , 2015.