A SURVEY OF BIG DATA ANALYTICS

MAYUR GANPAT SAKPAL, SREELAKSHMI NAIR

Abstract—

In the information era, enormous quantity of data have become present on hand to decision makers. Big data mention to datasets that are not only big, but also high in variation and speed, which makes them hard to handle using conventional tools and techniques. Due to the speedy growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. Further, decision makers need to be able to gain valuable insights from such diverse and fast changing data, ranging from daily deal to customer interactions and social network data. Such value can be provided using big data analytics, which is the application of developed analytics techniques on big data. This paper aims to analyze some of the different analytics methods and tools which can be applied to big data, as well as the chance provided by the application of big data analytics in various conclusion domains.

Index Terms: big data,methods,results,analytics

INTRODUCTION:

Big data analytics is the often a conglomerate process of examining large and varied data sets, or big data, to uncover information -- such as hidden patterns, unknown correlations, market trends and customer preferences -- that can help organizations make informed business decisions. Big Data is a term used to express a collection of data that is huge in size and yet growing exponentially with time, which means that data to be handled is so large and complex that none of the traditional data management tools are able to store it or process it efficiently. The main goal of big data analytics is to help companies make more informed business decisions by authorize data scientists, predictive model, and other big data analytics professionals to analyze large volumes of transactional data, as well as other forms of data that may be untapped by more conventional Business That could include Web server logs and Internet click stream data , social media content, contents from customer emails and survey response Big data can be analyzed with the software tools commonly used as part of advanced analytics disciplines such as anticipating analytics, data mining, text analytics and statistical analysis. Mainstream BI software and data visualization tools can also play a important role in the analysis process. But the semi-structured and unstructured data may not fit well in conventional data warehouses based on relational databases. Additionally, data warehouses may not be able to handle the processing demands posed by sets of big data that need to be updated regularly Those technologies form the core of an open source software framework that supports the processing of large and varied data sets across clustered systems.

- *Mayur Sakpal is currently pursuing masters degree program in Information Technology in Mumbai University, India, PH-7506775960 E-mail: mgsakpal21@gmail.com*
- *Sreelakshmi Nair is currently a lecturer in Mumbai University, India, PH-8169119781. E-mail:sreelakshmi@mes.ac.in*

METHOD:

In terms of method, big data analytics dissimilar importance from the common statistical approach of test design. Analytics starts with data. Ordinarily we model the data in a way to explain a response. The purpose of this approach is to predict the response behaviour or understand how the input variables relate to a response. Normally in statistical experimental designs, an experiment is developed and data is recover as a result. This allows to generate data in a way that can be used by a statistical model, where certain conclusion hold such as independence, routine, and unsystematic. In big data analytics, we are presented with the data. We cannot design an

experiment that satisfy our favourite statistical model. In large-scale applications of analytics, a large amount of work (normally 80% of the effort) is needed just for cleanse the data, so it can be used by a machine learning model. We don't have a single technique to follow in real large-scale applications. Normally once the business problem is defined, a analysis stage is free from design the methodology to be used. However general guidelines are applicable to be mentioned and apply to almost all problems. One of the most important tasks in big data analytics is statistical mode ling, meaning supervised and unsupervised grouping or backsliding problems. Once the data is cleaned and pre processed, available for mode ling, care should be taken in evaluating different models with reasonable loss metrics and then once the model is implemented, further evaluation and results should be reported. A common pitfall in predictive mode ling is to just implement the model and never measure its performance.

R**ESULTS**:

publications from the first paper recognize in this study, from 1996 until today. While there were smaller than four publications from 1996 to 2012, it should be noted that there has been a steady grow in relevant publications starting in 2013 and peaking at25 published papers in 2016. Since then, the number of papers published has grow less. It is noteworthy that the proportion of papers dealing with very specific product-related maintenance has getup . These papers often address specifics of maintenance that are related to the design of the product, rather than the actual nurture or data analytics-based maintenance of the corresponding produce processes. These papers have therefore been filtrate out, leading to a decrease in the overall number of papers. It should also be mentioned that some of the 2017 contributions may not have been fully indexed at the time the databases were searched.

DISCUSSION:

The first applicable publication dated back to 1996. Even
though the term "big data" was introduced early 2000, the paper of 1996 can be categorize, under big data from today's view with reference to the technology at that time and the approach made. The number of publications was in comparison low, with a maximum of three per year until 2012. The number of publications per year increase from 2013 onwards and peaked at 25 in 2016. It should be noted that only papers up to December 2017 were considered (indexed up to February 2018). It is likely that this number will change during the year. Overall, it can be seen that research efforts in big data analytics in the field of maintenance within production or manufacturing is growing, year on The three Vs of big data could be found in most papers, starting with the heavy volume of data used as the basis for big data analytics. We initial intended to classify the existing writings on an ordinal scale for each of the three Vs. One problem that arose during extraction of the papers was that very few authors give concrete numbers on the amount of data and data sources they had considered when writing their papers. Volume form.: the rationale :supporting analysing the volume of data was to quantify how much was being used to develop analytical models. In extra words, how much data was required for the models to work and how much training data for all types of learning algorithms were available. Only a few writer give details in their present some instance, the volume of historical data. For example, one benefaction stated that they used the previous 60 days of data to build a predictive model. speed aspect: the idea behind mapping research efforts to this attribute of big data was to get an idea of how the analytical models cope with the speed of incoming data (data collecting, is a common technique used to handle the velocity aspect, among other things). We hoped that best-practice solutions could be identified in order to give clear hints to other research teams. As with the volume aspect, most authors gave very few details on this aspect of their work. This aspect needs to be looked further by reviewers of those contributions, since validation and verification of results is made impossible

CONCLUSION:

The aim of this research was a literaturebased evaluation of the application of big data analytics in maintenance. so seven research questions were formulated, in context of which the reports-based research was conducted. A systematic approach was chosen for the research process, including a search and strain strategy. The resulting papers were grade based on defined arrangement. and then envision Based on these categorised papers, it was possible to answer the research questions defined at the beginning. In summary, there is an visualize trend towards applying big data analytics to maintenance. It is noticeable that different data analytics methods and technologies were used in combinations, even though predictive analytics appears the most popular in recent years. character. and machine learning are the most commonly applied when it cover technologies. Alongside a general trend to word sampling big data analytics to maintenance, a activity.

trend can be observed.

REFERENCES

1 . Adams, M.N.: Perspectives on Data Mining. International Journal of Market Research 52(1), 11–19 (2010)

2. Ahuja , I.P.S.; Khamba , J.S. An evaluation of TPM implementation initiatives in an Indian manufacturing enterprise. J. Qual. Maintenance . Eng. 2007, 13, 338–352. [Cross Ref]

3. Cooke, F.L. Plant maintenance strategy: Evidence from four British manufacturing firms. J. Qual. Maintenance. Eng. 2003, 9, 239–249. [Cross Ref]

4. Fraser, K.; Hvolby, H.-H.; Tseng, B. Maintenance management models: A study of the published literature to identify empirical evidence. A greater practical focus is needed. Int. J. Qual. Relia. Manage. 2015, 32, 635–664. [Cross Ref]

5. Cooke, F.L. Implementing TPM in plant maintenance: Some organizational barriers'. Int. J. Qual. Reliab. Manage. 2000, 17, 1003–1016 5. Bokrantz, J.; Skoogh, A.; Berlin, C.; Stahre, J. Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030. Int. J. Prod. Econ. 2017, 191, 154–169. [Cross Ref]

6. Fraser, K.; Hvolby, H.-H.; Watanabe, C. A review of the three most popular maintenance systems: How well is the energy sector represented? Int. J. Glob. Energy Issues 2011, 35, 287–309. [Cross Ref]

7. Ahuja, I.P.S.; Khamba, J.S. An evaluation of TPM initiatives in Indian industry for enhanced manufacturing per performance. Int. J. Qual. Reliab. Manag. 2008, 25, 147–172. [Cross Ref]

formance. Int. J. Qual. Reliab. Manag. 2008, 25, 147–172. [Cross Ref]

8. Subramaniyan, M.; Skoogh, A.; Salomonsson, H.; Bangalore, P.; Gopalakrishnan, M.; Sheikh Muhammad, A. Data-driven algorithm for throughput bottleneck analysis of production systems.