A Study on Relationship between Data Mining and Big Data

Kavita Srivastava

Faculty of BCA, Department of Business Management and Entrepreneurship, Dr. R. M. L. A. University Ayodhya, India

Abstract— Big data is an expression for a data set. Big data sets are those that exceed the straightforward sort of database and data taking care of models that were utilized in before times, when big data was progressively costly and less achievable. Experts also initiate the distinctiveness and function of several popular running platforms. In this paper, we elaborate to identify the challenges and issues of big data and data Ming with closed relationship. We recognized quite a lot of factors from the big data and data Ming perspective and we also decorated the data Ming issue that justify considerable additional research and development. However, database and data taking care of models issues there a crucial difficulty for user to get used to into data Mining.

Keywords— Mining, Architecture, Challenges, Big Data, Research Issues.

I. INTRODUCTION

Data mining suggests the activity of encountering huge data sets to scan for essential or suitable data. This sort of activity is incredibly a certified instance of the commonplace saying "looking for a needle in a parcel." The thought is that associations accumulate huge arrangements of data that may be homogeneous or subsequently assembled [8]. A researcher expects access to smaller, progressively unequivocal bits of data from those huge sets. They use data mining to uncover the pieces of data that will enlighten expert and help diagram the course for a business [6]. Data mining can incorporate the use of different kinds of programming groups, for instance, examination instruments. It will in general be robotized, or it might be, as it were, work genuine, where particular masters send unequivocal inquiries for data to a narrative or database [4, 10]. All things considered, data mining implies undertakings that incorporate reasonably complex chase exercises that landing concentrated on and unequivocal results. Data mining is used for researching and examining considerable proportions of data to find structures for big data. The presence of big data, the data mining is dynamically unavoidable. Four or five quite a while back, associations assembled all data of trade set away in a lone database. Today, volume of data is assembled have exposed. Promoters can moreover assemble data about each exchange people are having about their picture [2]. It requires the execution of new systems, advancement and organization frameworks that are overall being insinuated to as big data. Today, big data is a big business. We can describe big data is a methodology that empowers associations to remove a data from tremendous

proportion of data. Big data is used data mining techniques since size of data is bigger [9].

II. BIG DATA AND DATA MING RELATIONSHIP

It's immense in nature in light of the fact that, there is the gathering of data from different sources together [5]. In the event that we consider the case of Face book, heaps of quantities of individuals are transferring their data in different sorts, for example, content, pictures or recordings [3]. The general population likewise keeps their data evolving consistently [1]. This enormous and immediately, time to time changing supply of the data is put away in a distribution center. This substantial stockpiling of data requires extensive region for real usage. As the size is excessively expansive, nobody is competent to control it oneself [7]. The Big Data should be constrained by isolating it in gatherings.
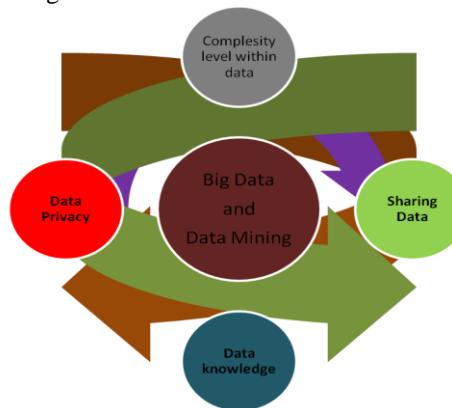

Figure 1 Big data and data mining Architecture

III. BIG DATA FEATURES

Big Data is put away at better places and furthermore the data volumes may get expanded as the data keeps on expanding persistently. In this way, to gather every one of the data put away at better places is that much costly. Assume, on the off chance that we use these common data mining techniques (those strategies which are utilized for mining the little scale data in our own PC frameworks) for mining of Big Data, and after that it would turn into a hindrance for it. Since the regular strategies are expected data to be stacked in principle memory, however we have overly expansive primary memory. There are three sectors at which the challenges for Big Data arrive.
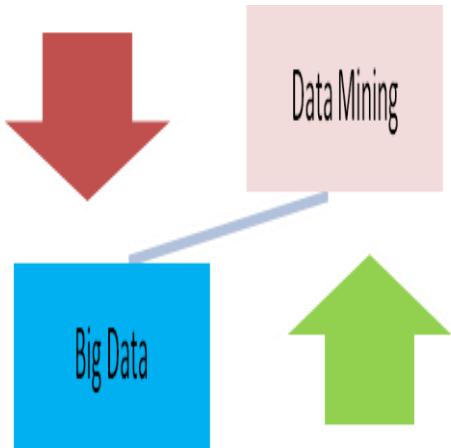
Figure 2 Balancing Relationship

IV. CONCLUSION

Big data must be a important for any large scale work. As such no methodology has been available in the literature that associated to the big data and data mining. big data and data mining is parallel to reduced methodology for evaluation of data set. Big data is a complete result of data mining which contributes in measure the fault. Big Data is becoming the new Final Frontier for scientific data research and for business applications. This research paper concludes that how will give a view to show the effect of Big data with data mining.

REFERENCES

[1]. K. Su, H. Huang, X. Wu, and S. Zhang, "A Logical Framework for Identifying Quality Knowledge from Different Data Sources," Decision Support Systems, vol. 42, no. 3, pp. 1673-1683, 2006.

[2]. E.Y. Chang, H. Bai, and K. Zhu, "Parallel Algorithms for Mining Large-Scale Rich-Media Data," Proc. 17th ACM Int'l Conf. Multimedia, (MM '09,) pp. 917-918, 2009.

[3]. D. Howe et al., "Big Data: The Future of Biocuration," Nature, vol. 455, pp. 47-50, Sept. 2008.

[4]. A. Labrinidis and H. Jagadish, "Challenges and Opportunities with Big Data," Proc. VLDB Endowment, vol. 5, no. 12, 2032-2033, 2012.

[5]. Chen et al. 2004, R. Chen, K. Sivakumar, and H. Kargupta, Collective Mining of Bayesian Networks from Distributed Heterogeneous Data, Knowledge and Information Systems, 6(2):164-187, 2004.

[6]. Chen et al. 2012, Yi-Cheng Chen, Wen-Chih Peng, Suh-Yin Lee, Efficient algorithms for influence maximization in social networks, Knowledge and Information Systems, December 2012, Volume 33, Issue 3, pp 577-601

[7]. Chu et al., 2006, Chu C.T., Kim S.K., Lin Y.A., Yu Y., Bradski G.R., Ng A.Y., Olukotun K., Map- reduce for machine learning on multicore, In: Proceedings of the 20th Annual Conference on Neural Information Processing Systems (NIPS '06), MIT Press, 2006, pp. 281-288.

[8]. Cormode G. and Srivastava D. 2009, Anonymized Data: Generation, Models, Usage, in Proc. of SIGMOD, 2009. pp. 1015-1018.

[9]. Brijesh Kumar Bhardwaj, "Performance Analysis and Evaluation in Data Mining: An Educational Perspective", International Journal of Scientific Research and Reviews, Vol. 7, Issue 4, July-Sept. 2018.

[10]. Anshul Mishra, Devendra Agarwal and M. H. Khan, "Availability Estimation Model: Fault Perspective", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 6, June 2017.