

# Leveraging Data-Centric Tools and Techniques in Application of Big Data and Predictive Analytics

Savar Sharma

*PDM University, Bahadurgarh, Haryana*

*<sup>1</sup>Received: 12 Oct. 2023; Accepted: 28 Dec. 2023; Published: 06 Jan. 2024*

---

## ABSTRACT

Big Data has created a fundamental locale of eagerness for study and research among experts and academicians. Movement in advancement is making it fiscally conceivable to store and examine gigantic proportions of information. Enormous Data consolidates a mix of composed, semi-sorted out and unstructured progressing information starting from a combination of sources. Farsighted Analytics provides a method of tapping learning from broad information files. Various visionary associations, for instance, Google, Amazon, etc have comprehended the ability of Big Data and Analytics to expand the upper hand. These techniques give a couple of chances like discovering precedents or better improvement figures. Directing and breaking down. Big data similarly sets up a couple of troubles - specifically estimate, quality, steadfast quality and satisfaction of data. This paper gives an expansive review of composing on Big Data and Predictive Analytics. It gives unpretentious components of urgent thoughts in this rising field. Finally, we have completed with disclosures of our examination and structured future research orientation in this field.

## INTRODUCTION

In the progressing years, Big Data starts from grouping of data sources in different structures. The volume, collection and speed of this data present amazing challenges for those directing server ranches. Coincidentally, preparing, stockpiling and examination capacities have compensated for lost time to address these troubles. The limit of far-reaching datasets has ended up being basic and temperate. Nearby traditional business data, firms are recognizing a motivating force from electronic long-range informal communication data obtained from areas.

## RESEARCH OBJECTIVES

Motivated by the requirement to likewise research the activity of big information and Predictive Analytics, this paper exhibits to interface the learning opening by achieving the following goals:

- a) To explore the best-in-class possibilities and challenges of Big Data and Predictive Analytics
  - b) To recognize openings in existing investigation and perceive further look into headings at work of Big Data and Predictive Analytics.
1. Research the ebb and flow composing the crucial contemplations of Big Data and Predictive Analytics
  2. Clear up the enhancement and ramifications of Big Data and Predictive Analytics

## RESEARCH METHODOLOGY

In the wake of describing the focuses of our investigation, we perceived catchphrases, for instance, " Big Data", " Big Data Analytics", "Insightful Analytics", "Online life Analytics" and "Twitter Analytics" for looking for study papers in best journals. The investigation papers were browsed year 1992 to 2016, with a bigger piece of them conveyed after year 2013. Frankly, over portion of the papers have been circulated over the last 3 years, which is an average marker of the significance of this subject (red shading in Figure 1 underneath).

---

<sup>1</sup> How to cite the article: Sharma S. (2024); Leveraging Data-Centric Tools and Techniques in Application of Big Data and Predictive Analytics; International Journal of Technology, Science and Engineering; Vol 7 Issue 1; 1-7

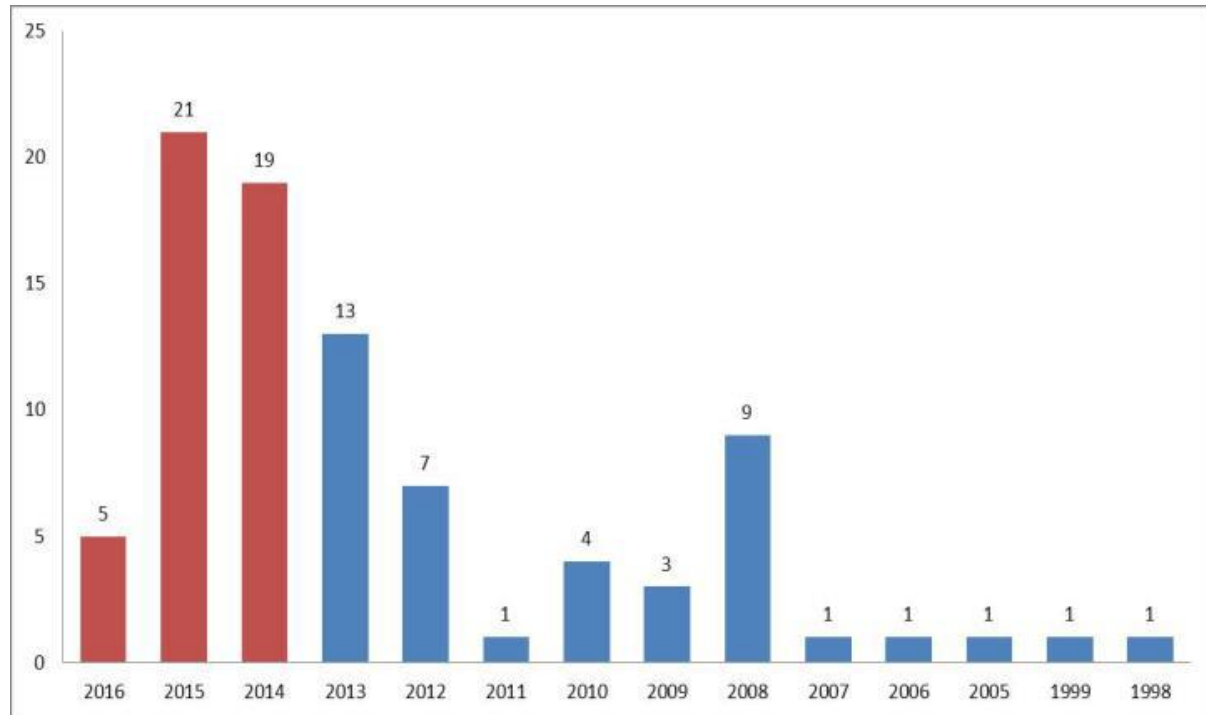


Figure 1. Big Data and Predictive Analytics of Year-wise Classification

### EVALUATION OF INFORMATION SYSTEMS LEADING TO BIG DATA

Over the past many years, information structures and the web have been big engaging impacts of globalization. From starting the use of information structures for sensible applications and departmental information systems, We have accomplished a period of "PDAs" and "Internet of Things". Essential clarification behind before time uses of data structures was record keeping and profitable preparing of business exchanges. Beginning now and into the not-so-distant, several achievements in programming designing and building have incited data change in the most recent couple of decades. Chen et al. (2012) outline this into 3 irrefutable stages as depicted in Table 1 underneath.

Table 1: Phases of Data Evolution Chen et al., (2012)

Phase	Description
I – Till year 1999	frameworks of database to save, break down and information created information in RDBMS
II– 2000- 2010	structures Wide utilization of Internet, segment and enhancement of web firms twitter, Google and so on., online business applications, electronic business, supply chains
III – 2010-	onwards Entry of bleeding edge PDAs, RFID, Sensor drives

Right when differentiated and customary data, Big Data changes in size and also in its casing. It gets included relentlessly instead of for the most part static data in the inheritance or ERP systems (swan, 2016).

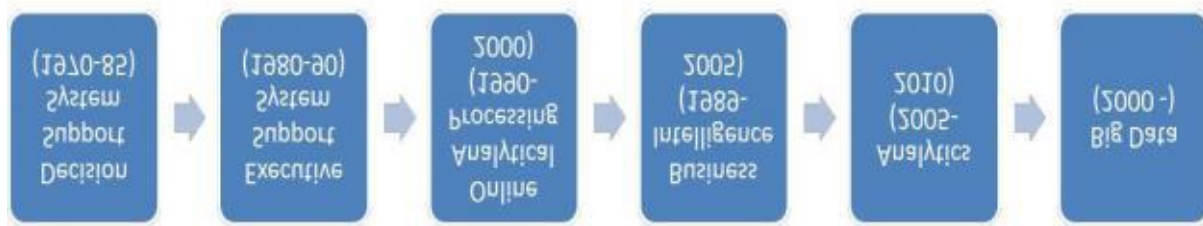


Figure 2. ERP frameworks

## BIG DATA

Fan et al., (2014) consider it an impact of open information. Big Data can't be described just by the volume of data, yet it fuses fast, unique grouping, intensive in augmentation, and social in nature (Kitchin, 2014). To say it doubtlessly, Big Data insinuates datasets. ( Waller and Fawcett 2013). With bytes of land and petabytes of data made in a constrained ability to centre a couple of hours

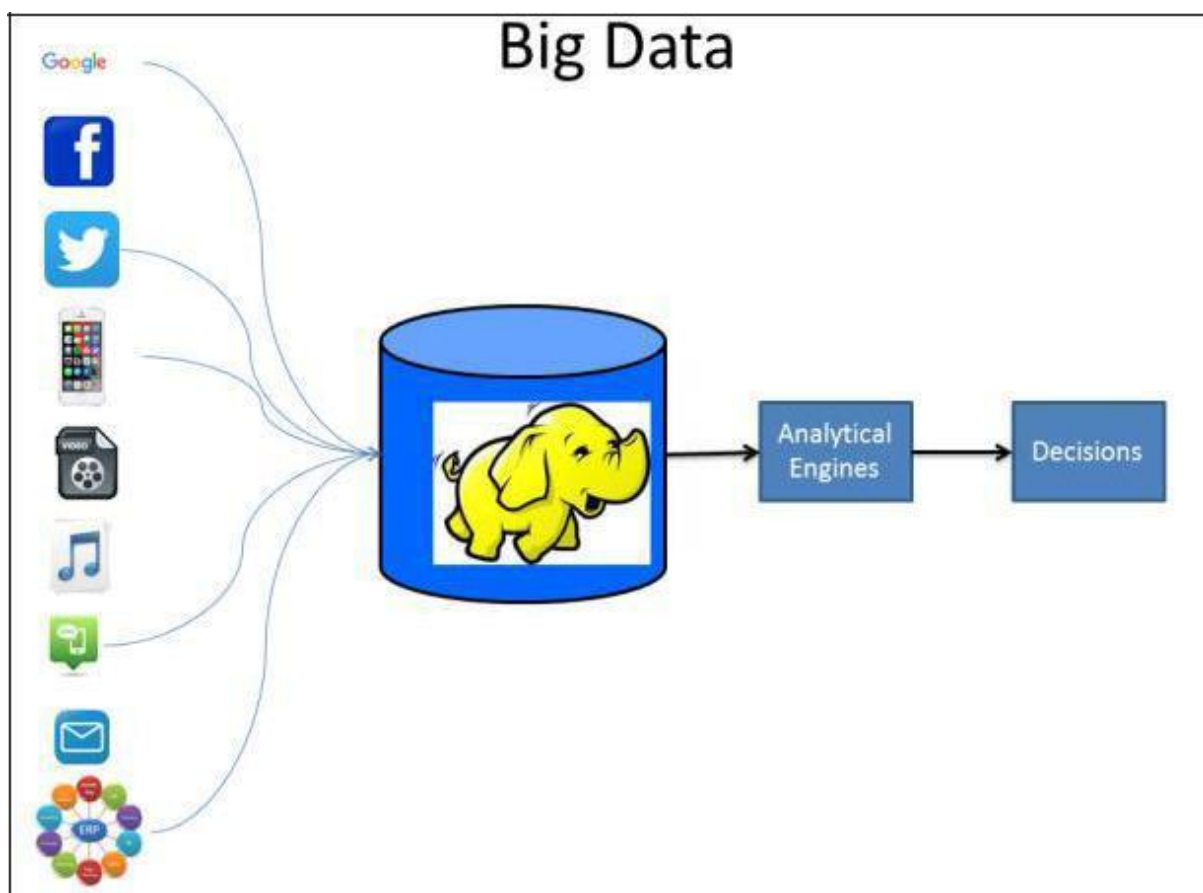


Figure 3: Overview of Big Data and Analytics

## DIFFERENT DATA TYPES

All the various sources create distinctive types of information that can be comprehensively delegated Structured, Un-organized and Semi-organized information. (Figure 4)

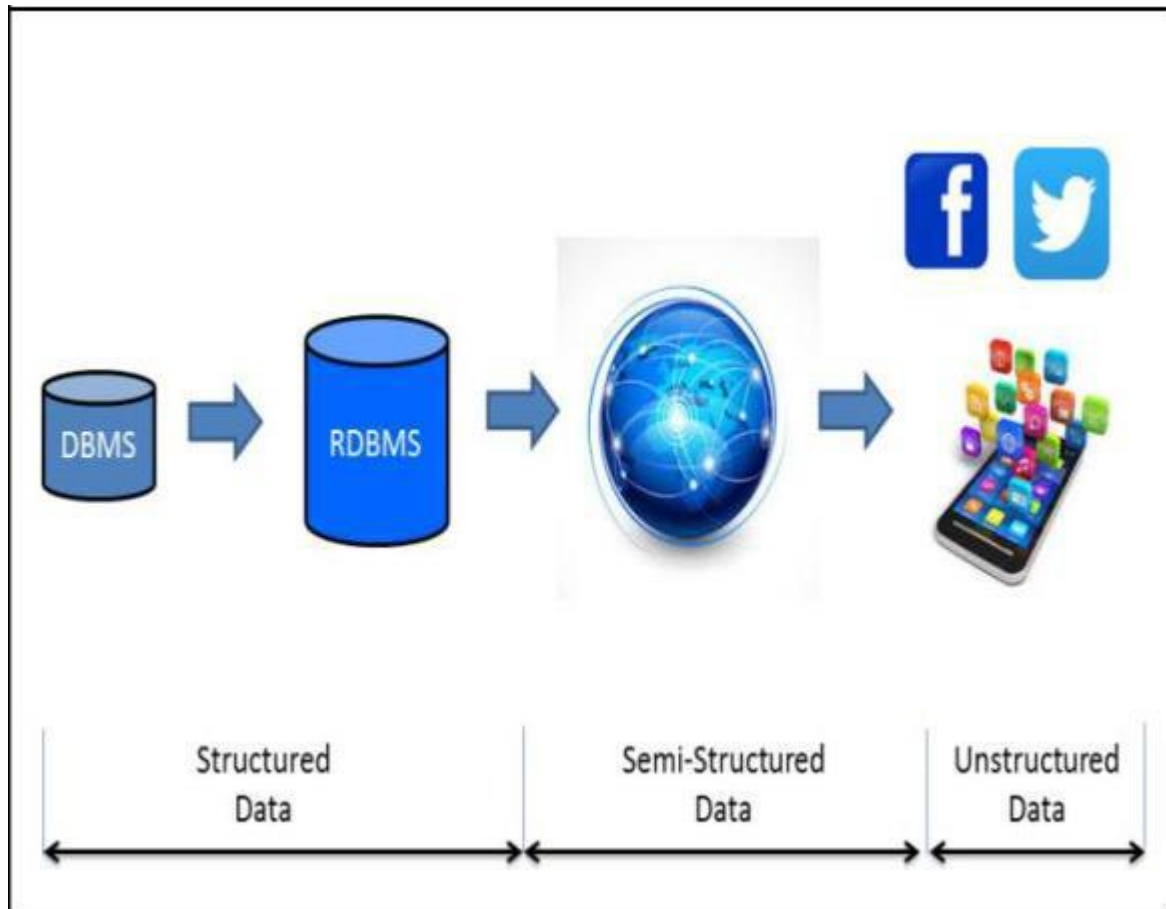


Figure 4: Data types in Big Data

**Structured Data:** The source of organized information is organizational data frameworks, for example, the purpose of sales data, batch forms, ERP frameworks, and broadened venture frameworks.

**Unstructured Data:** Unstructured data begins from an assortment of sources, for example, online networking, text messages, messages, connections, recordings, pictures and sound documents.

**Semi-structured data:** This begins with a wide range of basis combinations of organized and formless information. Different data frameworks have been used for velocity, proficiency and exactness of data trade with partners. Rigid utilize messages for correspondence, RFID innovation for quicker preparation in coordination (Deng et al., 2010) or sensor gadgets for following items.

### CHARACTERISTICS OF BIG DATA

There are two characteristics of big data are following as-

**Volume:** Big Data volume is extraordinary data –Bytes of Tera or Petabytes of data get gathered in the range of a few hours in business or online life databases. The extent of data is expanding like 40 months (Davenport, 2014).

**Velocity:** The velocity of data gathering is at noteworthy rates in both customary endeavour systems and in electronic life.

### BIG DATA ANALYTICS

Big Data Analytics has its foundations in the prior information investigation approaches utilizing quantifiable techniques, for example, descend into sin, factor examination, and so forth. It combines data mining from speed data streams and sensor information to get advancing examination (Chaw et al., 2015), (Sharma and Kitchin, 2016), Pawlak et al., (2011).

## 1) PREDICTIVE ANALYTICS

Perceptive Analytics is described as the path toward finding imperative instances of data using plan affirmation frameworks. In like manner, insinuated as Advanced Analytics, it just means utilization of data examination strategies to answer locations or deal with issues (Bose, 2014), (van der Aalst, 2012). The Hadoop framework gives a response for dealing with these examination essentials. In light of the source and nature of different data, there are diverse examination systems that support data mining and measurable investigation techniques. Content Analytics systems gather ceaseless and critical information from unstructured data sources, for instance, records, messages, site pages and online life.

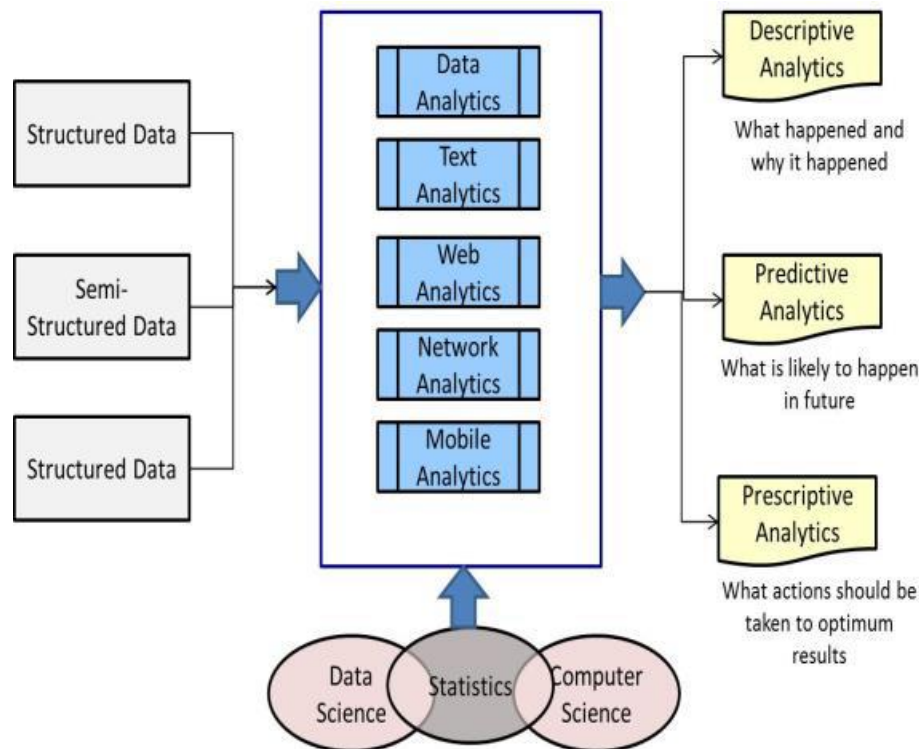


Figure 5: Framework for Predictive Analytics

## 2) SOCIAL MEDIA ANALYTICS

Internet-based life examination is a creating thought which is ending up some segment of standard advancing strategy. It depends on online systems administration media data made on areas. It is stressed over making and surveying informatics instruments and structures to accumulate, screen, examine, layout and envision online networking information (Dubey et al., 2011). Enormous Data made on the internet containing texts, tunes, pictures, accounts, etc. People share information through texts, chronicles, pictures, tunes, etc. They frequently express their objective to purchase a thing, request analysis, and share their organization experience or thing reviews on the web-based life. This data from electronic life has noteworthy information for firms that can separate and mine this data. Figure 6 below abbreviates the non-explicit process purchasers seek for procuring things. This technique diagrams how clients investigate things using web records and electronic life. They moreover contribute their viewpoints, and sentiments utilizing online systems administration media while acquiring the thing. Advancement enables firms to use this medium to help business methods and key fundamental initiatives. We condense a part of the habits by which online life examination can be valuable for the associations in the accompanying sub-region.



Figure 6. Content generation on social media

### 1. Sentiment Analysis:

A couple of associations accumulate and analyse client emotions imparted through online systems administration media about their things or organizations. There are a couple of various methods for examination using basic tongue planning, end word references or machine learning figuring in a wide grouping of undertakings. Goals, for instance, [www.sentiment140.com](http://www.sentiment140.com) give bits of information on customer estimation on things or organizations. Google Analytics gives diverse organisations – progressively shrewd publicising (data-driven appropriate and practical advancing), significant customer encounters.

### 2. Competitive Intelligence:

In the present very focused markets, business leaders look for purchaser criticism about their very own items and additionally those of their opponents. Kim et al., (2016) recommend the utilization of online networking examination for social affair-focused knowledge about the association's items and items offered by rivals in a similar market fragment. Their near examination of Twitter information for 2 contending advanced cells over some undefined time frame uncovered a relationship between assumptions communicated on the web-based life and a contrast in market interest for the 2 items.

## IMPLEMENTATION OF ANALYTICS

Like any data framework experience, executing Big Data Analytics experience inside an affiliation has its great difficulties. Setting up the required foundation, higher introductory cost, changes to business shapes (Bose, 2014) and receptiveness of experienced information masters – these are a piece of the inconveniences in the execution of Analytics inside an alliance. To use bits of gaining from Big Data, industry and the adroit system will require experienced masters in Data Science and Predictive Analytics. This field requires both space learning and a wide strategy of quantitative aptitudes.

## CONCLUSION

Big data is depicted by its volume, speed and assortment as it is made through formed and unstructured sources. It may be moreover gathered into 3 sub courses of action in light of the direction behind the examination. Clever examination regulates envisioning or envisioning future results in the context of mining existing information. Information can be sourced from business or electronic life. In any case, what is essential is to isolate it and get a comprehension of "what can occur". This capacity to imagine gives enormous capacity to prepare for conflict for business firms. It can give a window of chance for game-plan ahead of schedule to governments in conditions, as example, ocean tempests or the spread of scourges.

## REFERENCES

1. M. K.Kakhani, S. Kakhani and S. R.Biradar, Research issues in big data analytics, International Journal of Application or Innovation in Engineering & Management, 2(8) (2015), pp.228-232.
2. A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, 35(2) (2015), pp.137-144.



3. C. Lynch, Big data: How do your data grow?, *Nature*, 455 (2008), pp.28-29.
4. X. Jin, B. W.Wah, X. Cheng and Y. Wang, Significance and challenges of big data research, *Big Data Research*, 2(2) (2015), pp.59-64.
5. R. Kitchin, Big Data, new epistemologies and paradigm shifts, *Big Data Society*, 1(1) (2014), pp.1-12.
6. C. L. Philip, Q. Chen and C. Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on big data, *Information Sciences*, 275 (2014), pp.314-347.
7. K. Kambatla, G. Kollias, V. Kumar and A. Gram, Trends in big data analytics, *Journal of Parallel and Distributed Computing*, 74(7) (2014), pp.2561-2573.
8. S. Del. Rio, V. Lopez, J. M. Bentez and F. Herrera, On the use of mapreduce for imbalanced big data using random forest, *Information Sciences*, 285 (2014), pp.112-137.
9. MH. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki and D. K. Grunwell, Health big data analytics: current perspectives, challenges and potential solutions, *International Journal of Big Data Intelligence*, 1 (2014), pp.114-126.
10. R. Nambiar, A. Sethi, R. Bhardwaj and R. Vargheese, A look at challenges and opportunities of big data analytics in healthcare, *IEEE International Conference on Big Data*, 2013, pp.17-22.
11. Z. Huang, A fast clustering algorithm to cluster very large categorical data sets in data mining, *SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, 1997.
12. T. K. Das and P. M. Kumar, Big data analytics: A framework for unstructured data analysis, *International Journal of Engineering and Technology*, 5(1) (2013), pp.153-156.
13. T. K. Das, D. P. Acharjya and M. R. Patra, Opinion mining about a product by analyzing public tweets in twitter, *International Conference on Computer Communication and Informatics*, 2014.
14. L. A. Zadeh, Fuzzy sets, *Information and Control*, 8 (1965), pp.338- 353.
15. Z. Pawlak, Rough sets, *International Journal of Computer Information Science*, 11 (1982), pp.341-356.
16. D. Molodtsov, Soft set theory first results, *Computers and Mathematics with Applications*, 37(4/5) (1999), pp.19-31.
17. J. F. Peters, Near sets. General theory about nearness of objects, *Applied Mathematical Sciences*, 1(53) (2007), pp.2609-2629.
18. R. Wille, Formal concept analysis as mathematical theory of concept and concept hierarchies, *Lecture Notes in Artificial Intelligence*, 3626 (2005), pp.1-33.
19. I. T.Jolliffe, *Principal Component Analysis*, Springer, New York, 2002.
20. O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis and K. Taha, Efficient machine learning for big data: A review, *Big Data Research*, 2(3) (2015), pp.87-93.
21. Changwon. Y, Luis. Ramirez and Juan. Liuzzi, Big data analysis using modern statistical and machine learning methods in medicine, *International Neurology Journal*, 18 (2014), pp.50-57.
22. P. Singh and B. Suri, Quality assessment of data using statistical and machine learning methods. L. C.Jain, H. S.Behera, J. K.Mandal and D. P.Mohapatra (eds.), *Computational Intelligence in Data Mining*, 2 (2014), pp. 89-97
23. A. Jacobs, The pathologies of big data, *Communications of the ACM*, 52(8) (2009), pp.36-44.
24. H. Zhu, Z. Xu and Y. Huang, Research on the security technology of big data information, *International Conference on Information Technology and Management Innovation*, 2015, pp.1041-1044.
25. Z. Hongjun, H. Wenning, H. Dengchao and M. Yuxing, Survey of research on information security in big data, *Congresso da society*.