

Insights of Mathematics for Big Data

D. N. Punith Kumar, Akram Pasha



Abstract: *Computer Science can be considered as one of the extensions made to the pure mathematical sciences that exhibit the design and development of many mathematical models to solve various engineering problems. Data storage and data processing are the two major operations that are primarily focused by any computational model while solving a problem. Mathematical modelling has been helped in producing the various computational models across several problems that are found in the field of computer science. Among many problems that are found in the area of computer science, data science and big data have recently geared up to solve many business oriented problems that are purely based on data analytics to enhance the profit by taking critical business decisions. Data Scientists and mathematicians are found to have a skeptical understanding or too little collaboration either in knowing the mathematical concepts behind big data technologies, or too little knowledge of applications of mathematical concepts in applications of big data, respectively. Therefore, in this paper, an effort is made to bring out the major mathematical concepts that have contributed in fueling the solutions for big data problems. The authors hypothesize that the work proposed in this paper would benefit any data scientist or a mathematician to clearly understand the bridge between the math and its application in big data analytics. The authors identify the mathematical concepts and their roles played while solving various tasks that are encountered in the domains of big data. Further, such an endeavor is expected to open up many opportunities for both mathematicians and big data professionals to work collaboratively, while encouraging and contributing in enhancing interdisciplinary research across many domains of engineering.*

Index Terms: *Big Data, Data Analytics, PCA, SVD, Laplacian Graph, Eigen values, Eigen Vectors, Linear Algebra.*

I. INTRODUCTION

In the modern era of digital world where we live, there is a generation of mammoth data by innovative technologies and old technologies. This mammoth data is termed as 'big data', one of the greatest approaches with the use of distributed technologies to provide solutions for any form of massive data. Big data is generated through major sources such as Internet, smart phones, management systems, e-commerce, healthcare, astronomy, and so on [1][2][3]. We are in the age of 4G and expected to be in 5G platform soon which will again increase the data generation and usage massively. Recently, IoT has entered into the market minimizing human interference producing more complexity in terms of data production and transmissions [4][5][6]. In all

of the aforementioned cases, issues and challenges arise in handling, analyzing data. Mathematics being the mother of computer science has its roots in solving every computational problem and helps solving big data issues in the most systematic and ethical manner [7][8][9]. The challenges are the problems on analysis of data including its storage, transmission, data noise handling, unstructured formats and dynamic nature of its generation. It is evident from the literature, that the solutions for these problems are provided by mathematical sciences allowing us to model by providing mega chest of various tools, methodologies and allowing us to validate these methodologies [10][11][12]. Over the period, a lot of mathematical research happened in big data. When we try to understand, experiment, model and control the systems involving huge data, it is necessary to understand the mathematics and science behind it. These understandings may vary which provokes towards automating the process which in turn relies on mathematical algorithms [13]. Big data comprises of mostly unstructured data having many types of data, such as images, audio, videos, text, and so on. Though, statistical methods have been providing necessary aid for analyzing data. Mathematics based algorithms have been playing a vital role and have got more importance. This paper attempts to provide a basic understanding and role of mathematical facets towards handling big data problems for any data scientist, data analyst or any computer science professional wishing to enter the field of big data analyst. The major contributions of this paper are that it:

- Introduces eigenvalues, eigenvectors, principal component analysis (PCA), the graph Laplacian, and singular value decomposition (SVD), the key mathematical concepts in big data analytics
- Presents the brief survey of applications of Eigen values and eigenvectors to investigate prototypical problems of ranking big data
- Presents the brief survey of applications of the graph Laplacian to investigate prototypical problems of clustering big data
- Presents the brief survey of applications of PCA and SVD to investigate prototypical problems of big data compression

The rest of the paper is organized in the following structure. Section-II introduces the concept of big data and its basic challenges. Section-III presents the key mathematical concepts such as eigenvalues, eigenvectors, PCA the graph Laplacian, and singular value decomposition (SVD). Section-IV presents the conclusion and future scope of the work carried out.

Manuscript published on 30 May 2019.

* Correspondence Author (s)

D.N.Punith Kumar, REVA University, Bengaluru, India.

Akram Pasha REVA University, Bengaluru, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

II. BIG DATA AND ITS CHALLENGES

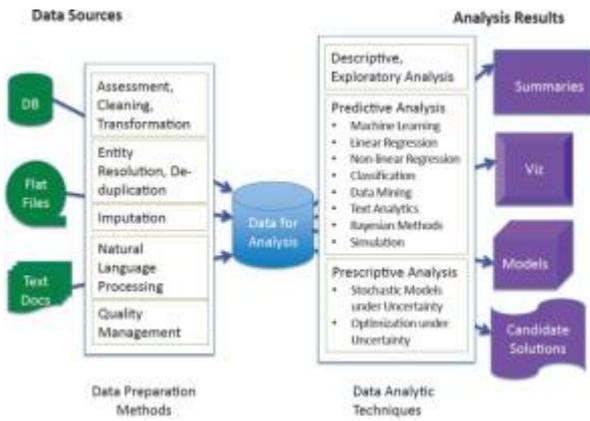


Fig 1: Characteristics of big data

a) Nature of Big Data:

A field that provides different ways to analyze and deal with large, complex data using traditional software is big data [5][6][7]. This term was popularized mainly by and was redefined in [8], it is the voluminous data that becomes difficult for traditional methods to handle. Big data gets its definition due to the nature of its characteristics. The characteristics of big data are its volume of existence, the variety of data formats that it produces, and velocity of its arrival and veracity of the data sources. Figure 1 shows the characteristics of big data [6]. Analysis of data leads to applications in various fields.

The data involved in the analysis is unstructured, semi-structured and structured. But the main concentration lies on unstructured. The processing of data can be done through algorithms and analytics. Methodologies to analyse data are namely machine learning, A/B testing, natural language processing. Technologies involving big data are numerous but not limited to business intelligence, cloud computing, IoT and web intelligence.

Figure 2 shows the stepwise process of data evolution and analysis results. The data prepared is analyzed through various techniques, then modeled and solution is provided [7][8].

Fig 2: Step wise process of data evolution and analysis

b) Applications of Big Data:

There is a rapid increase in the demand of specialists in the field of information management due to big data applications. Many top notch companies have spent billions on specializing in data management and analytics [9]. Big data also increases the efficiency in terms of innovation, productions and cost for many government processes. Figure 3 shows one the methods of capturing and storing big data in the real time applications requiring rapid data analytics. Advanced analysis creates opportunities which are cost effective to enhance decision making strategies in the sectors like health care, crime, security etc. The analysis of big data has contributed enormously in healthcare sector by providing analysis on medicines, prescription, prediction of risk in clinical trials, patient reports, etc which still requires some improvement [10]. The usage of big data technologies is more in recent trends of media which aims to serve on mindset of customers. Big data also helps in improving training competitors using sensors [11]. This will help to predict winners in a match. Therefore more the application in various sectors provides more exposure and challenges to be faced, analyzed and developed for the future digital era.

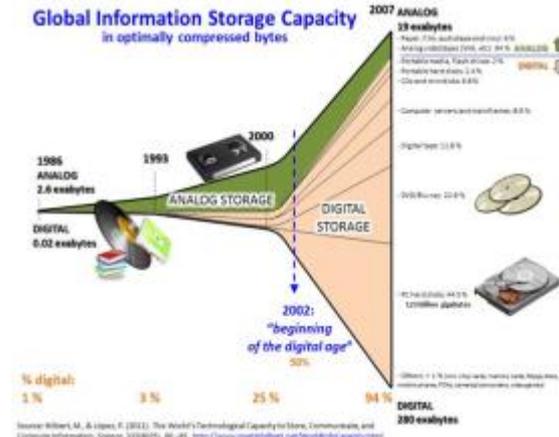


Fig 3: Diagrammatic representation of storage capacity and global digitization [6].

c) Major Challenges in Big Data:

Managing big data using traditional methods is a challenge to the digital community because new innovations, technologies produce huger, complex data which are faster and variable [14]. To provide feasible solutions for these challenges one must know complete cycle of the analysis. In this cycle, data is to be managed and analyzed. Managing data involves acquiring, recording, extracting, cleaning, integrating, aggregating and presenting [15]. The analysis includes modeling, analyzing, interpreting, and drawing conclusions. The other challenges are to understand the unique features of big data like high dimensionality, security, scale, privacy and so on. In order to address these challenges, a strong knowledge on properties related to mathematical sciences is a must.



The mathematical sciences community should train the current generation for the upcoming challenges in analytics domain [16]. A strong knowledge on numerical linear algebra, topological data analysis, matrix decompositions, tensor decompositions, graph theory, and network theory would help in solving issues existing in future [17].

III. KEY MATHEMATICAL CONCEPTS IN BIG DATA ANALYTICS

Mathematical sciences have been contributing to analyze big data but there is still plenty of room for the development and the contribution to be made in this domain. Among many directions of the works that could be made in the directions mentioned, the contributions can be made by providing various mathematical tools that are innovatively designed; to analyse and manage huge complex data, to design and develop inference systems to conclude on noisy and large data, and to develop mathematical algorithms from the major sub-fields of mathematics such as linear algebra, and to perform computations for large sizes of data through various mathematical representations or data structures for data [18][19]. The list of mathematical methods mostly used in the analysis of big data, to name some are Linear Algebra, Probability theory and statistics, Multivariate calculus, Algorithms and complexity and the lists goes that would mostly represent any data facet mathematically.

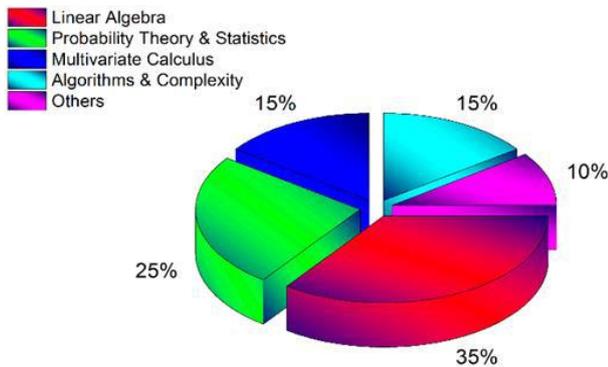


Fig 4: Graphical representation of Mathematical approaches for big data [16]

Figure 4 represents the importance of mathematics in big data that is most recently found in the literature [20]. The topics in linear algebra such as matrix operations, vector spaces, Eigen values, Eigen vectors, orthogonalization and orthonormalization are required to understand as they map with the concepts of optimization techniques used in machine learning [20], the tool that is widely used in the field of knowledge discovery processes, such as data mining. The topics in probability and statistics like combinatorics, theoretical distributions, and sampling methods are required to understand as they are mapped with the concepts of feature selection and feature extraction used in the domain of big data or data mining. The topics like differential, integral calculus, vector calculus, data structures, dynamic

programming, and randomized algorithm are required to understand as they are mapped with the key concepts related to the computational efficiency of any algorithm studied to discover its space and time complexity [21].

A.EIGEN VALUES AND EIGEN VECTORS FOR BIG DATA RANKING

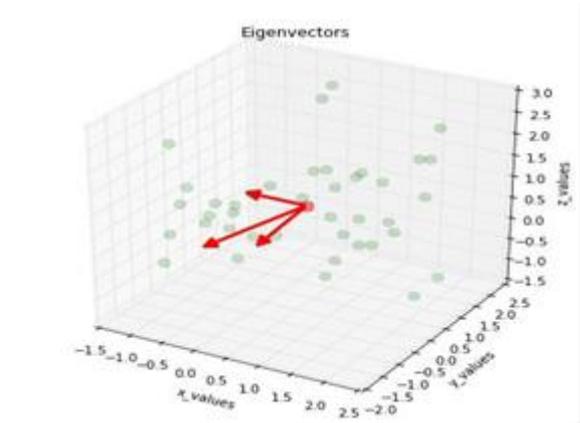


Fig 5: Eigen values and Eigen vectors

The figure 5 shows the typical example of Eigen values and Eigen vectors. The work of [22] finds the Eigen values and Eigen vectors of a particular system that is perturbed using known values. This can be illustrated by considering the solutions to generalized Eigen value problem using the equation (1).

$$LX_{0i} = \lambda_{0i} N_0 X_{0i} \tag{1}$$

In equation (1), L_0 and X_0 are matrices. Suppose we know the Eigen values λ_{0i} and Eigen vectors X_{0i} , and we want to change the matrices then the new values can be found using equation:

$$LX_i = \lambda_i NX_i \tag{2}$$

In equation (2), $L = L_0 + \delta L$, $N = N_0 + \delta N$ and δL and δN are perturbed terms. Then the perturbed Eigen values and Eigen vectors will be of the form shown in the equation (3):

$$\lambda_i = \lambda_{0i} + \delta \lambda_i \quad \text{and} \quad X_i = X_{0i} + \delta X_i \tag{3}$$

Applications:

Eigen decomposition is a method to find Eigen values and Eigen vectors from a decomposed matrix. This method is applied in machine learning such as PCA where these values help to find significant features of massive data. In recent years, global energy interconnection has become hot topic and big data serves as one of its technologies. There are many advantages and applications of big data in support of interconnection involving application of Eigen values in the process [24]. Eigen values helps in obtaining asymptotic results to estimate factor models under certain constraints of rank [25]. Various network models can be classified using a reduction scheme of linear algebra.

The Eigen values are used to find the rank and signature of the matrices involved in the process [26]. The new challenges of processing massive data by models in machine learning can be done using decomposition concepts of matrices showing the importance of Eigen values in the process [27]. Before the conduction of statistical inference, it is necessary to apply dimension reduction techniques in the case of complex data. In this process principal component analysis becomes an efficient tool using Eigen values in the process of analyzing multivariate data focusing on ranking and other parameters [28]. Spatial filtering using Eigen vectors is demanding approach applied in various studies by researchers and practitioners for the analysis of large samples and to calculate its low rank approximations [29]. The sum of Eigen values and Eigen vectors involved in process of predicting ratings or ranks which are not known in the field of graph signal processing can be reduced by developing new designs [30]. The literature shows that the concepts of Eigen values and Eigen vectors are the key concepts in mathematics that makes the foundation for many algorithms of machine learning that are used to tackle many problems in the domain of big data.

B. GRAPH LAPLACIAN AND BIG DATA CLUSTERING

The Laplacian matrix is a matrix representation of graph describing properties of a graph [20] [22]. This has wide applications in machine learning. This matrix is as defined in the equation (4).

$$L = D - A \tag{4}$$

The computation of degree matrix is defined with the following property:

$$D[i][j] = \text{degree of node } i \text{ if } i = j \text{ and } 0 \text{ if } i \neq j$$

The computation of adjacency matrix is defined with the property:

$$A[i][j] = 1 \text{ if there is an edge between node } i \text{ and node } j; \text{ otherwise } 0$$

Applications:

Graph Laplacian is used to compute least significant Eigen vectors and dimensions of clusters in spectral clustering. Representing complex data in the case of pattern recognition is a big challenge. Drawing correspondence between Graph Laplacian and other methodologies can develop algorithms to represent the data of high dimensions [26]. The interpretation of signal processing can be done using Laplacian matrices and presented in the form of Fourier transform for graph signals [27]. During the construction of graphs in spectral analysis, using the right content of the data is important. Many a times this data is incomplete and this issue is addressed by distilled graph clustering method performed using graphs of Laplacian matrices. The resulting graph is

used for spectral clustering on social network containing large data [31]. During the process of removing noise and errors from the data, a robust graph learning scheme can be applied developed using Laplacian matrix theory. This scheme improves the performance of data clustering and recovery [32]. Adaptive graph learning methods also use graph Laplacian techniques to improve the performance of clustering. The newly developed algorithms improve the performance of clustering methods on benchmark data [33]. Matrix theory plays a vital role in data clustering and it is evident from the literature that clustering algorithms based on properties of graph Laplacian are more efficient compared to other algorithms that are not using Laplacian matrix [34].

C. PCA & SVD IN BIG DATA COMPRESSION

PCA:

It is a statistical procedure that uses an orthogonal transformation to convert a set of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components [27].

Applications:

PCA minimizes the information loss under certain models. It is used to study compression techniques of smart meter big data through smart grid. This helps in the evaluation of compression methods for data compression [35].

PCA is used in developing compression algorithms to improve performance of compression system with error resiliency [36]. It is also demonstrated that along with PCA, set partitioning hierarchical trees can be used to compress images. A model is developed to use the benefits of these methods in the field of bio medical research [37]. One of the realistic applications of PCA is in compression of wind tunnel data performed by National Institute of Standards and Technology [38].

SVD:

SVD is factorization of a real or complex matrix. In this method of linear algebra, a real or complex matrix is decomposed into unitary and diagonal matrix [23].

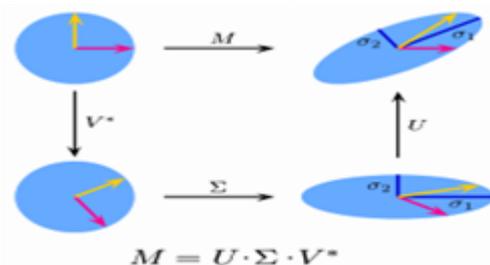


Fig 7 : Interpretation of SVD

Figure 7 presents diagrammatic structure of SVD where M is an m x n matrix, U is an m x m unitary matrix and V is an n x n matrix.

The upper left disc shows canonical unit vectors, upper right disc shows transformation as singular values, lower left disc shows rotation and lower right disc shows scaling in vertical and horizontal directions.

Applications:

The major application of SVD is in the field of cloud computing where K-cloud SVD is studied in collaboration with union of subspaces [31].

SVD reduces the chances of data loss occurring during compression of data [32]. SVD also helps in processing huge data with tolerated elapsed time [33]. We can also achieve scalable performance using the combination of R and SVD to distribute series code to a parallel code [34]. SVD is also used to compress ECG data providing quality control encrypted data [39]. In the field of signal processing, the matrix is decomposed by SVD giving singular values yielding signals with reduced noise [40]. An algorithm developed using SVD and ASCII character encoding compressing MMioSigs can be seen in the work of [42].

D. MATHEMATICAL MODELLING AND BIG DATA

Mathematical modelling is developing models describing a system using mathematical concepts and language [25]. These models can be dynamical, statistical, differential or game models. Algorithms based on these models yield better performance during dynamic data handling under various circumstances [26]. Also mathematical models play vital role in data mining. Mathematical models of big data can help various sectors to meet their challenges and can also be extended to various other domains. The mathematical models used by researchers are inspired by exposure in mathematics [30]. Mixture of models has drawn attention of many researchers aiming at analysis of data clustering [42]. Many models result in effective revealing of structures of clustering [43]. Mathematical based models have importance in scientific simulations estimating and extrapolating the reduction ration helping scientists to take decisions on data compression [44]. Mathematical models become key element in development of multigrid algorithms from graph Laplacian systems [45].

IV. CONCLUSION AND FUTURE SCOPE

The collaboration of mathematical concepts with big data benefits the professionals from both mathematics and data analytics domain. Such an effort would enhance the quality of the solutions obtained in the domain of data analytics and would also promote the essence of mathematical modelling. This paper attempted to present the influence and importance of mathematical concepts in the domain of big data. An attempt was made to illustrate the fusion of mathematical approaches with technologies for data analysis. However, though, various techniques and tools from mathematical sciences domain are available in today's world for analysis of big data, a having a strong background of pure mathematical

concepts behind those techniques and tools would certainly produce high quality solutions to all the domains of data analytics. There is plenty of room for more interdisciplinary research and development with various domains of engineering with mathematical sciences to enhance the efficiency of current technologies. Exploring the influence of probabilistic concepts and randomness in inference drawing in the domain of data analytics shall be carried out in the future work.

REFERENCES

1. Sun, Zhaohao, and Paul P. Wang. "A mathematical foundation of big data." *New Mathematics and Natural Computation* 13.02 (2017): 83-99.
2. Laval, Philippe B. "Introduction to the Mathematics of Big Data." Retrieved September 4 (2015): 2016.
3. Sun, Zhaohao, and P. Wang. "Big data, analytics and intelligence: an editorial perspective." *Journal of New Mathematics and Natural Computation* 13.2 (2017): 75-81.
4. Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, <https://doi.org/10.1177/2053951716631130>.
5. Kitchin, Rob, and Gavin McArdle. "The diverse nature of big data." *SSRN Electron. J* (2015).
6. Kitchin, Rob. "Big data—Hype or revolution." *The SAGE handbook of social media research methods* (2017): 27-39.
7. Ibrahim Abaker Targio Hashem Albrar Yaqooba Nor Badrul Anuara Salimah Mokhtara Abdullah Gania Samee Ullah Khanb. The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, Volume 47, January 2015, Pages 98-115.
8. Amir Gandomi *, Murtaza Haider, Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management* 35 (2015) 137–144.
9. Uthayasankar Sivarajah, Muhammad Mustafa Kamal, Zahra Irani Vishanth Weerakkody, Critical analysis of Big Data challenges and analytical methods, *Journal of Business Research*, Volume 70, January 2017, Pages 263-286
10. John R. Mashey (25 april 1998). "Big Data ... and the next wave of infrastress" (pdf). slides from invited talk. usenix. retrieved 28 september 2016.
11. <https://www.nap.edu/read/23670/chapter/6>.
12. Steve Lohr (1 February 2013). "The Origins of 'Big Data': An Etymological Detective Story". *The New York Times*. Retrieved 28 September 2016.
13. De Mauro, Andrea; Greco, Marco; Grimaldi, Michele (2016). "A Formal definition of Big Data based on its essential Features". *Library Review*. 65 (3): 122–135. doi:10.1108/LR-06-2015-0061.
14. Hilbert, Martin. "Big Data for Development: "A Review of Promises and Challenges. *Development Policy Review*". martin.hilbert.net. Retrieved 7 October 2015.
15. White Paper: Big Data for Development: Opportunities & Challenges (2012) – United Nations Global Pulse". [Unglobalpulse.org](http://unglobalpulse.org). Retrieved 13 April 2016.
16. Hilbert M. Big data for development: From information-to knowledge societies. Available at SSRN 2205145. 2013.
17. Elena Kvochko, Four Ways To talk About Big Data (Information Communication Technologies for Development Series)". Worldbank.org. 2012-12-04. Retrieved 30 May 2012.
18. Daniele Medri: Big data & Business: An ongoing revolution, statistical views, 21 october 2013.
19. Admire Moyo. "Data scientists predict Springbok defeat". www.itweb.co.za. Retrieved 12 December 2015.
20. Zhang, Dongxia, and Robert Caiming Qiu. "Research on big data applications in Global Energy Interconnection." *Global Energy Interconnection* 1.3 (2018): 352-357.
21. Kang, U., and Christos Faloutsos. "Big graph mining: algorithms and discoveries." *ACM SIGKDD Explorations Newsletter* 14.2 (2013): 29-36.

22. Belkin, Mikhail, and Partha Niyog. "Laplacian eigen maps for dimensionality reduction and data representation." *Neural computation* 15.6 (2003): 1373-1396.
23. Pavez, Eduardo, and Antonio Ortega. "Generalized Laplacian precision matrix estimation for graph signal processing." 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016.
24. Lei, Qi, Kai Zhong, and Inderjit S. Dhillon. "Coordinate-wise power method." *Advances in Neural Information Processing Systems*. 2016.
25. Bai, Jushan, and Serena Ng. "Rank regularized estimation of approximate factor models." *Journal of Econometrics* (2019).
26. Valdano, Eugenio, and Alex Arenas. "Exact rank-reduction of network models." *arXiv preprint arXiv: 1811.08299* (2018).
27. Muravev, Anton, et al. "Acceleration Approaches for Big Data Analysis." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.
28. Fan, Jianqing, et al. "Principal component analysis for big data." *Wiley StatsRef: Statistics Reference Online* (2018): 1-13..
29. Murakami, Daisuke, and Daniel A. Griffith. "Eigenvector spatial filtering for large data sets: fixed and random effects approach." *Geographical Analysis* 51.1 (2019): 23-49.
30. Huang, Weiyu, Antonio G. Marques, and Alejandro R. Ribeiro. "Rating prediction via graph signal processing." *IEEE Transactions on Signal Processing* 66.19 (2018): 5066-5081.
31. Liu, Wenhe, et al. "Learning Distilled Graph for Large-scale Social Network Data Clustering." *IEEE Transactions on Knowledge and Data Engineering* (2019).
32. Kang, Zhao, et al. "Robust Graph Learning from Noisy Data," *IEEE transactions on cybernetics* (2019).
33. Liu, Tianchi, et al. "An adaptive graph learning method based on dual data representations for clustering." *Pattern Recognition* 77 (2018): 126-139.
34. Alami, Réda, "Graph Signal Processing: Towards The Diffused Spectral Clustering." (2019).
35. Wen, Lulu, et al. "Compression of smart meter big data: A survey." *Renewable and Sustainable Energy Reviews* 91 (2018): 59-69.
36. Kwan, Chimman, and Yvonne Luk. "Hybrid sensor network data compression with error resiliency." *Data Compression Conference*. 2018.
37. Reddy, Rajasekhara M., et al. "A new approach for the image compression to the medical images using PCA-SPIHT." *Biomedical Research* (0970-938X) (2018).
38. Zhou, Jin, and Chimman Kwan. "A Hybrid Approach for Wind Tunnel Data Compression." *Data Compression Conference*. 2018.
39. [39] Liu, Ting Yu, Kuan Jen Lin, and Hsi Chun Wu. "ECG data encryption then compression using singular value decomposition." *IEEE journal of biomedical and health informatics* 22.3 (2018): 707-713.
41. Schanze, Thomas. "Compression and noise reduction of biomedical signals by singular value decomposition." *IFAC-PapersOnLine* 51.2 (2018): 361-366.
42. Windmann, Stefan, et al. "Big data analysis of manufacturing processes." *Journal of Physics: Conference Series*. Vol. 659. No. 1. IOP Publishing, 2015.
43. Mukhopadhyay, Sourav Kumar, M. Omair Ahmad, and M. N. S. Swamy. "SVD and ASCII Character Encoding-Based Compression of Multiple Biosignals for Remote Healthcare Systems." *IEEE transactions on biomedical circuits and systems* 12.1 (2018): 137-150.
44. McLachlan, Geoffrey J., Sharon X. Lee, and Suren I. Rathnayake. "Finite mixture models." *Annual review of statistics and its application* ,6 (2019): 355-378.
45. Wang, Pingxin, and Yiyu Yao. "CE3: a three-way clustering method based on mathematical morphology." *Knowledge-Based Systems* 155 (2018): 54-65.
46. Lu, Tao, et al. "Understanding and modeling lossy compression schemes on HPC scientific data." 2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS). IEEE, 2018.
47. Fox, Alyson, and Thomas Manteuffel. "Algebraic multigrid for directed graph Laplacian linear systems (NS-LAMG)." *Numerical Linear Algebra with Applications* 25.3 (2018): e2152.
48. Snijders, Chris, Uwe Matzat, and Ulf-Dietrich Reips. "" Big Data": big gaps of knowledge in the field of internet science." *International Journal of Internet Science* 7.1 (2012): 1-5.
49. https://en.wikipedia.org/wiki/Mathematical_model.
50. Vijaylakshmi, s. p. and j. Priyadarshini. "Big data analysis based on mathematical model: a comprehensive survey." (2015).
51. Dalla Vecchia, Rodrigo. "The relationship between Big Data and Mathematical Modeling: A discussion in a mathematical education scenario." *Themes in Science and Technology, Education* 8.2 (2016): 95-103.
52. Qiu, Junfei, et al. "A survey of machine learning for big data processing." *EURASIP Journal on Advances in Signal Processing* 2016.1 (2016): 67.
53. Sandryhaila, Aliaksei, and Jose MF Moura. "Big data analysis with signal processing on graphs." *IEEE Signal Processing Magazine* 31.5 (2014): 80-90.
54. Che, Dunren, Mejdil Safran, and Zhiyong Peng. "From big data to big data mining: challenges, issues, and opportunities." *International conference on database systems for advanced applications*. Springer, Berlin, Heidelberg, 2013.
55. Yu, Le, et al. "BC-PDM: data mining, social network analysis and text mining system based on cloud computing." *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012.
56. Joshi, Milan, and Dhananjay Joshi. "A survey of Topological Data Analysis Methods for Big Data in Healthcare Intelligence." *International Journal of Applied Engineering Research* 14.2 (2019): 584-588.
57. Nachtsheim, Abigail C., and John Stufken. "Comments on: Data science, big data and statistics." *TEST* (2019): 1-4.

AUTHORS PROFILE



Dr. Punith Kumar D.N. pursued Bachelor of Science from Bengaluru University, India in 2004, Master of Science from Bengaluru University in year 2007 and PhD from Visvesvaraya Technological University in 2018. He is currently working as Assistant Professor in School of Applied Sciences, Department of Mathematics, Reva University since 2012. He is a life member of ISTE since 2008. He has published research papers in reputed international journals. His main research work focuses on Biofluidynamics, Mathematical modeling, Big Data Analytics, IoT, Computational Fluid Dynamics, Probability and Statistics, Numerical Analysis and Partial differential equations based education. He has 12 years of teaching experience and 9 years of Research Experience.



Prof. Akram Pasha, Associate Professor, School of Computing and Information Technology, REVA University, Bengaluru, India owns Bachelor of Engineering from Mysore University, Master of Engineering from Bangalore University and currently pursuing Ph. D from Visvesvaraya Technological University, Belgaum, India. His major areas of research interests are applications of Computational Intelligence techniques on Data analytics.