A Collective Study of Machine Learning (ML) Algorithms with Big Data Analytics (BDA) for Healthcare Analytics (HcA)

Pradeep K R [1], Dr.Naveen N C [2]

[1]*Asst professor, Dept. of CSE, K.S. Institute of Technology, Bengaluru,* [2] *Professor, Dept. of CSE, JSS Academy of Technical Education, Bengaluru,*

Abstract: *Big Data Analytics (BDA) is one of the rising innovations as it guarantees to give better gain of knowledge from gigantic and heterogeneous information. BDA includes choosing the appropriate Big Data (BD) storage and calculated structure expanded by adaptable Machine Learning (ML) techniques. In spite of the huge buzz around BDA and its focal points to new problems, novel approaches of data capture, storage, analysis and visualization are accountable in favor of the materialization of the BDA research field. ML algorithms can be used in BDA to construct finer and extra precise inferences. Though the challenges BD imposes, these algorithms necessitate by adapting and optimizing to precise applications. This research work focuses and describes different ML algorithms in BDA that are useful in Health care Analytics (HcA).*

Keywords-Machine Learning, Big Data, Big Data nalytics, , Healthcare Analytics.
.

I. INTRODUCTION

Highlight ML is an area of gaining knowledge which provides computers the capability to study exclusive of being programmed, aiming to understand computational mechanisms in which experience can lead to improved performance [1]. A great number of ML algorithms have been proposed in the last decades, such as neural network, decision tree, support vector machine-nearest-neighbor, genetic algorithms, Q-learning, etc. this have been used in various domains such as natural language processing ,pattern recognition, autonomous control and systems robotics, [2, 3]. The goal of ML is to facilitate a method to learn from the history or current data and utilize that data to formulate predictions or else decisions concerning unidentified outlook actions, at the compassion of ML is the data to power the models, and the novel epoch of BD is to catapult ML to the vanguard of industry and research applications.

BD not only accounts for the large volume of data, but also for the unstructured nature of that data, known as variety, and the need of immediate processing of this data, known as velocity. The propagation of BD has enforced us to move around not only with data processing frameworks, but also implementations of ML algorithms. This is the main reason why BDA has drawn so much attention from academicians and researchers where sophisticated logical techniques work on BD sets. Thus, BDA is really about two things one is BD and the other is analytics in addition to that need to know how these two encompass up to craft solitary one of the majority thoughtful trends in today's technology. However, despite the tremendous interest in BD analytics and its advantages, an extensive literature survey focused on parallel data-intensive ML algorithms for BD has not been conducted so far. This has encouraged us to provide a comprehensive survey of distributed parallel ML algorithms for BD along with various optimization and performance metrics for evaluation. The task of BDA is divided into two phases. The first phase involves setting the stage for BD processing and includes the selection of BD storage and BD computational framework. The second phase involves the designing and implementation of ML tools for extracting insights from the data.

Health care Analytics (HcA) is an expression known towards characterizing the healthcare analysis actions which can remain initiated as per consequence of collected data related to four fields in healthcare, such as patient performance and sentiment data, clinical data, pharmaceutical and R&D data along with claims and cost data.

II. LITERATURE SURVEY

Having the complete significance of large volume of BD, traditional analytics tools were incorrect towards capture of data. With the same context if the volume of data stands huge, comprehensive analysis close through a succession for probable correspondences plus associations among different data is too immense for any analyst to check every hypothesis and to obtain every significance of hidden data [5]. Online analytical processing is just a systematized annex of the essential analytics which still rely on an entity to express actions. For this ML is perfectly intended for accomplishing the accessibility hidden within BD. ML hand over's on the guarantee of extracting importance from big and distinct data sources through outlying less dependence scheduled on individual track as it is data determined and spurts at machine scale. It is fine suitable towards the intricacy of handling

through dissimilar data origin and the vast range of variables as well as amount of data concerned where ML prospers on increasing datasets. The extra data supply into a ML structure, the more it be able to be trained and concern the consequences to superior value of insights. At the liberty from the confines of individual level thought and study, ML is clever to find out and show the patterns hidden in the data.

BD not only leads to the method used for predictive analytics, it also advances to the unified technique of view towards knowledge extraction and interpretation. Conventionally, data science is dependably been overwhelmed towards experimentation analysis, an methodology that gets to be unthinkable at large datasets would be expansive which is more heterogeneous. Amusingly, [6] accessibility from claiming fewer choices in constructing predictive models, in light of not many tools are considered in handling big datasets in a sensible amount of time . Furthermore, traditional statistical results naturally emphasis on static analytics which remains restricted towards analysis of tests solidified in time, which often brings about surpassed and questionable conclusions. ML approach is able to resolve the problems in applications with a set of common methods which vary as of more usual statistical approach [7]. This importance stays happening for actual and well accessible predictive analytics, with completely programmed and basic methodologies shorten a few from the characteristic data responsibilities, ML is ruling its method towards exacting a suitable meaning out of it, based on it better decisions will be made. With this ML is the science of training systems to learn from data and output appropriate response without being explicitly programmed [8].

The Persistence of Big Data Analytics: Data analytics is a way of acquiring knowledge from many origins aimed at extra use and instance decision making with predictions [9]. Correlated towards conventional data analytics which has enormous, assorted and numerous origins add the feature of BD that provides deceit set of tests to accomplish the objective of knowledge discovery and data mining. To confront by above said objection, researchers encompass projected paths which includes:

1. Effectiveness enhancement of lone-resource knowledge discovery approach [10],
2. Effective data-mining methods [10]
3. Dual Information Distance (DID) technique intended for development of decision trees [11],
4. Indication of Non-Obvious association Analysis [12] ,
5. Maximal-association imperatives for the most part related pattern extraction [13],
6. Data-driven axes conception model for correlation capacity [14],
7. Extensible I/O-bound parallel accession prejudice tumble for BDA [15]
8. Service-oriented advancement for reasoning function [16]

Blackett [17] defined three levels of data analysis such as predictive analytics, prescriptive analytics and descriptive analytics where the majority of the approaches above can be grouped.

The most critical period of the BD assessment series remains with data analysis that stresses by means to take out useful data which suggest inferences with keeping decision making. The kind of analysis trouble restricts the record of relevant data analysis techniques to the dilemma. By the expression analysis dilemma means, whether it is a classification difficulty for the purpose of estimation like time sequence prediction, a clustering difficulty, if it is about decision dependencies or relations etc. The next group of requirements is disturbed by means of preferences concerning the explanation [18].

Data analytics always aims to extort constructive advice as of the focus that is to be embattled beneath thoughtfulness, a small amount of the possible functions are listed below:

In the direction of generalizing and interpreting the information,

1. En route to propose and help decision-making
2. To analyze and determine purposes behind responsibility
3. To ensure if the data are authentic
4. To foresee future event

The following chart provides the assistance that will assures implementation in various type of BDA surveyed by TDWI 2016 Research of fourth quarter. [19]
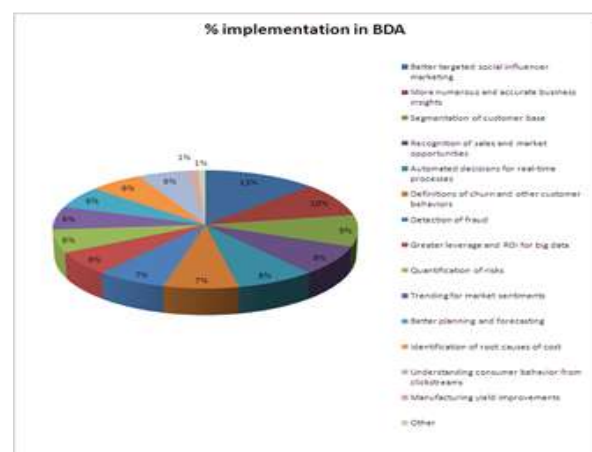

Figure 1: Barriers to BDA the top potential barriers to implementing BDA

The following are the different types of analytics depending upon the ML with BDA.

1. Text analytics: [20] Text Analytics, correspondingly acknowledged as text mining, is a

method stating to obtain functional data along with facts as of unstructured text. Here text contains social media content, WebPages, e-mail content etc. which is dependably chosen on refined usefulness. Text mining remains as an integrative range by the intersection related to data mining, computational semantics, information retrieval, machine learning and statistics. And this analytics is mainly established on text depiction and Natural Language Processing (NLP) which will be able to improve the convenient information associated to text conditions.

2. Web Analytics: Web Analytics intents towards regaining, extracting and calculating data for knowledge discovery as of web forms and services involuntarily. This analytics is established on quite a lot of additional stages for example information retrieval databases, NLP and text mining. Here this analytics is classified into three extents of significance such as web usage mining, content mining and structure mining [20]. Web usage mining cites to mining subordinate data cause via web sessions and it builds data from web server access, proxy server and browser logs, user sessions, cookies, and user profile. Here web content mining remains to the disclosure of functional evidence on website content that includes data for example audio, text, video, hyperlinks and image. The detection of the form basic relationship structures arranged in the web is known as web structure mining. At this point, design signifies the grid of relations in a site.

3. Multimedia Analytics: Obtaining interested knowledge and understanding the semantics occupied in multimedia data is referred as Multimedia analytics. New area of discussion beneath multimedia analytics adds to multimedia explanation, multimedia definition, indexing plus betterment. Multimedia explanation references to accredit images along with videos including group of labels so as to portray their substances at syntactic ranks. Multimedia summarizations consist of audio and video summarization so that it abstracts and makes a share of data as of the unique insides. Multimedia indexing with betterment engages explanation, repository then grouping of multimedia information intended on behalf of acceptable as well speedy improvement.

4. Network Analytics: The basis of the network analysis comprises sociology network analysis [21] and bibliometric analysis [22] which is contradictory close to promising social network analysis owing towards swift expansion in networked social networks. Social networks have excessive extent with correlation and contented data; such that link data perform connection involving graph structures with individual and contented data consists of text, multimedia in the networks and images. Social networks will be capable of being anticipated by

means of graphs, now that the vertex shows a type, and the authority harmonizes towards union connecting the reciprocal persons.

5. Mobile Analytics:[57-59] Through the accelerated progression of mobile computing along with new mobile stations as well as operations with cell phones, sensors and RFID stand located worldwide. At present, mobile analysis is far as of older but is facing definite significant objections induced by the built-in uniqueness of mobile data, for instance redundancy prosperity, noisiness, mobile responsiveness, and action compassion. New approaches with mobile expertise and wireless devices must be driven locating to the body locating system networks for actual observing of personality healthiness.

6. Health Care Analytics (HcA): [9] Health data capacity is probable to flourish radically in due year's forward which leads to the techniques that influence BD efficiently. Here data includes individual and resident's data which notifies both medical doctor and patient through the decision-making method and helps to decide the mainly suitable healing option to a particular patient. With the help of digitization and combination effectively helps to use BDA in healthcare institutions compromising from individual-medical doctor hospital and multiple sources of groups to big hospital networks set to appreciate important assistance [23]. Significant assistance comprise finding diseases at former phases at what time they will be able to be treated further effortlessly and successfully, supervision of precise personality health and recognize health care fraud more rapidly with proficiently. Several queries related to Health Care Analytics can be addressed with help of BDA where it provides definite outcome that can be predicted and anticipated reliable on large amount of chronological data, such as duration of stay, patients preferring optional surgical procedure, patients who likely will not advantage as of surgical procedure, patients at danger for medical complexity, patients at risk of cancer survivability rate or diabetes evolution.

III. MACHINE LEARNING ADVANCES

A large set of ML algorithms are developed to build ML models and execute an ML process [24]. These algorithms can be classified on the basis of learning method which is as follows:

1. Regression algorithms: [26] Regression is concerned with design the association among variables i.e. iteratively advanced using a rate of error in the predictions done by the model. The task of predicting the value of a continuously varying variable is Regression in which it includes most popular algorithms such as Stepwise Regression, [24] Multivariate Adaptive Regression Splines (MARS),

Locally Estimated Scatterplot Smoothing (LOESS) ,Logistic Regression and Ordinary Least Squares Regression (OLSR)

2. Instance-based Algorithms (I-bA): It is a decision problem among occurrences of training data which are vital to the model. Such types of methods are usually assembled with the database for training data and evaluating test data towards the database by means of comparison evaluation, within sort to locate the greatest counterpart and build a prediction. These methods are also called lazy learner where they simply accumulate training data and waits until it is provided with a test data, then performs the learning. So lazy learner takes a reduced amount of time in training but extra time in predicting. The popular I-bA is Self-Organizing Map (SOM), Learning Vector Quantization (LVQ), k-Nearest Neighbour (kNN) and Locally Weighted Learning (LWL).

3. Decision Tree Algorithms: As a predictive model, decision tree learning benefits from a decision tree, which brings about the classification or regression rules by coercively, partitioning the dataset. For classification and regression problems, decision trees will be trained on data which will be used. With the categorical target values supported by the decision trees, they are identified as classification trees, however, where target values are constant; they are identified as regression trees. Decision trees subsist of a root, internal node or decision node, and leaf nodes representing class labels. Decision trees are frequently rapid and precise with an immense choice related to [machinelearningmastery.com] machine learning. Effectively main popular decision tree algorithms are Classification and Regression Tree (CART), Chi-squared Automatic Interaction Detection (CHAID), C4.5 and C5.0 (different versions of a powerful approach), Iterative Dichotomiser 3 (ID3), Conditional Decision Trees, M5 and Decision Stump.

4. Clustering Algorithms: Clustering is a technique of classification of objects into dissimilar groups. It partitions the data set into subsets or clusters, as a result the data in every subset allot a few regular attribute frequently, related to a few clear distances assess. [25]The best accepted clustering algorithms are k-Means, Hierarchical Clustering, k-Medians and Expectation Maximization (EM).

5. Artificial Neural Network Algorithms (ANN): ANN are models which uses supervised learning which are constructed based on the composition of biological neural networks. It has artificial neurons which are extremely weighted interconnections in the midst of units and learns by modification the association weights to execute parallel distributed processing. They are also called as parallel distributed processing networks which include algorithms such as [26] Back-Propagation, Perceptron, Radial Basis Function Network (RBFN) and Hopfield Network.

6. Deep Learning Algorithms: [26] Deep Learning approaches is an recent revise towards ANN which takes advantage of profuse cost-effective calculation. This algorithm is apprehensive through construction to a great extent of superior and additional complex neural networks, as numerous techniques are troubled by semi-supervised learning problems pointed to big datasets enclose extremely modest labeled data. The novel algorithms based on deep learning algorithms are Deep Belief Network (DBN), Restricted Boltzman Machine (RBM), Convolutional Neural Network (CNN) and Stacked Auto -encoders.

7. Ensemble Algorithms: Ensemble ways are methods based on unsupervised learning poised of various weak learner models which remain separately trained as of these predictions stand collective within various ways towards making overall prediction. It divides the training data keen on quantity of subsets of data to which autonomous learning models are constructed. All learning models are collective to build accurate hypothesis. The popular ensemble algorithms are AdaBoost, Bootstrapped Aggregation, Gradient Boosting Machines (GBM), Random Forest, Stacked Generalization (blending), Boosting, Bagging and Gradient Boosted Regression Trees (GBRT).

IV. MACHINE LEARNING IN HEALTHCARE ANALYTICS

ML is perfect for handling the favorable circumstances concealed in BD. It bears on the assurance of deriving value from big and unrelated data sources with quite less dependence on individual path [27]. As it is data driven ML runs at machine scale and also well-matched to the intricacy of handling the disparate data origins and the enormous choice of variables and data's are involved. On contrary with traditional analysis, ML prospers on expanding datasets. Any extra data given into a ML system makes more of it that can learn and pertains the outcome to superior feature observation. At liberty of initiation to the restrictions of individual scale view and analysis, ML is capable to learn and exhibit the patterns hidden in the data.

The purpose of data analytics technologies into healthcare has provoked significant benefits in mutually with data analysis and health area. ML technologies and predictive analytics encompass for decades in number of industries. This has impact on the healthcare area that has proceeded to adopt the technologies for a variety of applications such as

chronic disease supervision, conscription predictions and population healthiness threat evaluation.

HcA provides important insights for the health of a person based on collected data and background information. Of these extra data is collected and it is available for analysis, these intuitions become more precise for the value-based medicine that has enough time to endorse positive actions that improves the outcome of humankind. In addition, the operation of ML grants to increase the usefulness of existing approaches and recognize the treatments and interface behavior that yields best results for particular patients with precise circumstances. This customized approach to care is at the keystone to tailor medical decisions and practices for the particular patient based on each person's inherited composition, atmosphere and everyday life.

Identity of BD in developing HcA

The way to progress towards result which includes cost for evaluating BD in healthcare industry is under demand to maintain costs and get better outcomes; here BD is proving to be an important advantage. But, data only can't be indicator, to get all details about how data is analyzed to get smarter output is done with proper intervention and treatment options.

Healthcare has to run through the benefits by some of major developments in data sharing through health information exchanges; data management: data collection from electronic medical records; and enhanced data analysis by enterprise data warehouses and new analytical tools which includes the capability to track trends and patterns as of numerous sources. Improved HcA continues to expand advantage with health tactics, by better data convenience that enhances to the raise of insight desirable to deliver enhanced member outcomes, better management decisions and quality of care. To achieve this HcA provides the inputs in following way

1. Good assessment for targeting the right people: A health plan includes assorted groups of citizens who could be in some position down the health and fitness scale. Providing concern to the persons who requires it starts through analyzing numerous sources such as claims data to associate contributed data to health threat evaluation. With big populations, this evolves still more significant with potentially assisted interference as a practice to recover health and decline expenses.

2. Conveying at exact involvement at the exact time: [28] Finding individuals who exists as a patient is one extent in which BD is successful in analytics, further assuring the efficient involvement is required on behalf of identifying each individual patient and it is afforded once required. This remains an area in which the technical improvements are

united by analytics that are motivating with developments. Here the capacity to provide the precise information at the exact time decides progress, as individuals start on recognizing their individual liability, examine their healthiness and reveal relevant data with their family members. [29] It also coordinates across with the locations and sources, entire family members must obligate the equivalent data and effort towards the equivalent aim. Through all enhanced data, with the ease of access and exploration of data, methods will be able to convey quicker recognition of greater threat associates, advise extra suitable involvements and give data determined intensive care.

3. Adapt programs with the closing of loop: With additional knowledge proceeding to healthiness in addition to diseases the configurations of concern are obtainable, extra functional awareness forces for the programs to be more rapidly accustomed. Enhanced HcA progresses to the improved programs and the capability to produce new ones. The prospective to progress the outcomes includes an outlay from the scrutiny of BD in HcA with the challenges to be addressed. The challenges were categorized by the following categories such as Analytic tools, Change management Costs, data management, Education, Integration and leadership, and Process, talent and technology. The facing of these challenges for HcA implementation has problems associated with lack of framework and outdated data. The revolutionary of BDA infrastructure relies on frequent data standards that provide users with contact to analytics generated from real-time data sources which will be solution to create flexible and beneficial healthcare system. The potential advantages to be gained at a national level seem to outweigh the investments already being made in finding ways to more quickly field and stitch disparate pieces of data into a holistic tapestry of health and well-being at the individual and aggregate population levels.

V. CHALLENGES REGARDING ML WITH HCA

The following are the challenges regarding ML with HcA

1. Designing extensible and responsive computational architectures for ML suitable for HcA [30]

2. Capability to build, discover, recognize uniqueness of data ahead of applying ML algorithms and tools.

3. Capacity to achieve statistical predicate invention with Structured Prediction by means of intractable inference

4. Ability to generalize all across domains and

learn many levels of structuring the data with deep combination of learning and inference to map between representations.

Large-scale ML, [30] such as ensemble learning, natural language processing (NLP), association rule learning, large-scale recommender systems, all are still facing the scalability difficulties [31] and the key problems to facilitate the ML methods that are inappropriate for resolving the BD classification problems are

1. ML method is trained on a specified datasets cannot be appropriate for another dataset i.e. the classification might not be potent over different datasets.
2. ML technique is normally trained by a definite amount of class types and therefore a great range of class categories are set up with robustly increasing dataset that drives directly towards imprecise classification outcomes.
3. ML technique is based on a lone learning mission, so this technique is not appropriate aimed at current numerous learning tasks plus knowledge transfer needs of BDA [32].

Traditional algorithms in ML usually do not scale to BD .The key complexity lies in the midst of their memory restraint. Even though algorithms characteristically presume that training data samples subsist in main memory, BD does not qualify into it. A general method of learning from a huge dataset is data allocation. By changing batch training scheduled on the unique training dataset by alienated computations on the dispersed subsets, where we can train an different prediction model at a surrender of accurateness. On an additional approach is by means of online learning, within which memory usage does not rely on dataset capacity. Equally online learning and distributed learning are not enough for learning as of BD branch because of the following challenges:
1. If the data size is bigger i.e. to be flexible by any online or distributed learning. Sequential online learning on BD requires large time for training on a sole machine. On the extra hand, distributed learning through a big number of machines curtails down the attained effectiveness per machine and alters the overall performance. [33]
2. The next cause is that joining real-time training and prediction has not been considered. BD is utilized once being stored in distributed storage; as a result, the learning process too favors to effort in a batch method [34].

VI. CONCLUSION

BDA and ML technologies are two techniques which provide the successful application of ML in HcA with intelligent algorithms and rich data sets which resolve the essentials required for researchers and healthcare industry. By incorporating ML techniques into conventional health care mechanism, healthcare industry can provide enhanced place to tackle the distinctive requirements of entity patients and enthusiastically manage healthcare. BDA in HcA is developing to a talented area of catering vision in distinction to extremely big data sets and enlightening results as of decreasing expenditures. BDA has the prospective to make over the method of life sciences and healthcare organizations using refined technologies to put on insights with starting their medical and added data repositories to build knowledgeable decisions. Overall analytics allow organizations to examine and discover data to recognize associations, trends, and patterns to disclose insights, as collective by means of business perspective and generate knowledge. HcA with BDA and ML helps to conclude the most possible outcomes and be improved set for the revolution. As healthcare continues its move to value-based medicine, researcher's helps to attain the need of innovative technologies to supervise the health of the patient population, by improving the outcomes, and to maintain costs under control which will be true when caring for individuals with chronic conditions. HcA provides the kind of cleverness that makes ideal intellect in the approaching years. As BDA becomes more main stream with ML algorithms in health care industry with improvised tools and techniques healthcare industry can provide true benefits in life sciences and healthcare in a compliant and cost-effective manner.

REFERENCES

[1] T.M. Mitchell, Machine Learning, McGraw Hill, New York, 1997.
[2] X.W. Chen, and X Lin, Big data deep learning: challenges and perspectives, IEEE Access 2 (2014) 514-525.
[3] N Jones, Computer science: the learning machines, Nature 505 (2014) 146-148
[4] en.wikipedia.org/healthcareanalytics
[5] Singh, Jainendra. "Real time BIG data analytic: Security concern and challenges with Machine Learning algorithm", 2014 ,Conference on IT in Business Industry and Government (CSIBIG), 2014
[6] www.ngdata.com
[7] J. Fürnkranz et al., Foundations of Rule Learning, Cognitive Technologies, DOI 10.1007/978-3-540-75197-7 1, © Springer-Verlag Berlin Heidelberg 2012
[8] Phil Simon (March 18, 2013). Too Big to Ignore: The Business Case for Big Data. Wiley. p.89.ISBN 978-1-118-63817-0.
[9] Big data analytics in healthcare: promise and potential Wullianallur Raghupathi1 and Viju Raghupathi Health Information Science and Systems 2014, 2:3 http://www.hissjournal.com/content/2/1/3
[10] Chen, R., Sivakumar, K., & Kargupta, H. (2004). Collective mining of Bayesian networks from distributed heterogeneous data. Knowledge and Information Systems, 6, 164–187
[11] Ben-Gal, I., Dana, A., Shkolnik, N., & Singer, G. (2014). Efficient construction of decision trees by the dual information distance method. Quality Technology & Quantitative Management (QTQM), 11, 133–147.

[12] Kogge, P. M., & Bayliss, D. (2013). Comparative performance analysis of a Big Data NORA problem on a variety of Architectures. In Geoffrey, F., Waleed, W. S. (Eds.), In Collaboration technologies and systems (CTS), 2013
International Conference on (pp. 22–34). Washington, DC, USA: IEEE. Soysal, Ö. M. (2015). Association rule mining with mostly associated sequential patterns. Expert Systems with Applications, 42, 2582–2592.

[13] Nakanishi, T ,"A data-driven axes creation model for correlation measurement on big data analytics" Information Modelling and Knowledge Bases XXVI, 272, 308, 2014

[14] Qin, C., & Rusu, F. (2013). Scalable I/O-bound parallel incremental gradient descent for big data analytics in Glade. In Proceedings of the second workshop on data analytics in the cloud (pp. 16–20). New York, NY, USA: ACM.

[15] Cheptsov, A., & Koller, B. (2013). A service-oriented approach to facilitate big data analytics on theWeb. In Bathe, K.J., Topping, B.H.V. (Eds.), In Proceedings of the fourteenth international conference on civil, structural and environmental engineering computing. Stirlingshire, UK: Civil-Comp Press.

[16] https://tdwi.org/research/2016/07/best-practices-report-improving-data-preparation-for-business-analytics.aspx

[17] Blackett, G. (2013). Analytics network – O.R. & analytics [online].Retrievedfromhttps://www.theorsociety.com/Pages/SpecialInterest/AnalyticsNetwork_analytics.aspx

[18] [13]Shekhar, Himanshu, and Manoj Sharma. "A Framework for Big Data Analytics as a Scalable Systems."Special Conference Issue: National Conference on Cloud Computing & Big Data 72-82,IJANA,http://www.ijana.in/Special%20Issue/C14.pdf.

[19] S. K. Pal, V. Talwar, and P. Mitra, „„Web mining in soft computing framework: Relevance, state of the art and future directions,'' IEEE Trans. Neural Netw., vol. 13, no. 5, pp. 1163–1177, 2002 .

[20] D. J. Watts, Six Degrees: The Science of a Connected Age. New York, NY, USA: Norton, 2004.

[21] J. E. Hirsch, "An index to quantify an individual's scientific research output,'' Proc. Nat. Acad. Sci. United States Amer., vol. 102, no. 46, p. 16569, 2005.

[22] [55-57] H. Zhang, Z. Zhang, and H. Dai, „„Gossip-based information spreading in mobile networks,'' IEEE Trans. Wireless Commun., vol. 12, no. 11, pp. 5918– 5928, Nov. 2013.

[23] H. Zhang, Z. Zhang, and H. Dai, „„Mobile conductance and gossip-based information spreading in mobile networks,'' in IEEE Int. Symp. Inf. Theory Proc. (ISIT), Jul. 2013, pp. 824–828

[24] H. Zhang, Y. Huang, Z. Zhang, and H. Dai. (2014). Mobile conductance in sparse networks and mobility-connectivity tradeoff. in Proc. IEEE Int. Symp. Inf. Theory (ISIT) [Online]. Available: http://www4.ncsu.edu/?hdai/ISIT2014-HZ.pdf

[25] Burghard C. Big Data and Analytics Key to Accountable Care Success. 2012.

[26] Sandhya N. dhage, Charanjeet Kaur Raina "A review on Machine Learning Techniques" In: Proceedings of International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 4, Issue: 3 (pp. 395– 399) ISSN: 2321-8169

[27] www.mo-data.com

[28] www.machinelearningmastery.com

[29] http://www.skytree.net/machine-learning/why-do-machine-learning-big-data/

[30] David Edward Marcinko, Hope Rachel Hertico, "Financial Management Strategies for Hospitals and Healthcare Organizations: Tools, Techniques, Checklists and Case Studies", 5 September 2013, CRC Press.

[31] https://www.hpe.com/h20195/v2/GetPDF.aspx/4AA6-7132ENW.pdf

[32] IHTT, Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry.2013

[33] Chen, C. L. P. & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275(10), 314-347.

[34] Suthaharan, S. (2014). Big data classification: problems and challenges in network intrusion prediction with machine learning. Performance Evaluation Review, 41(4), 70-73.

[35] Prateek Bihani1 and S. T. Patil "A Comparative Study of Data Analysis Techniques " ,ijettcs Volume 3, Issue 2, March – April 2014

[36] LaValle S, Lesser E, Shockley R, Hopkins MS, Kruschwitz N. Big data, analytics and the path from insights to value. MIT Sloan Manag Rev. 2011; 52:20–32.

Author[s] brief Introduction

[1]**Pradeep K R** received her Master's Degree in Networking and internet engineering in 2010 from SJCE, Mysore affiliated to Visvesvaraya Technology University, Belagavi, India and is currently pursuing his PhD from Visvesvaraya Technology University, Belagavi, India. He is currently working as Assistant Professor, Department of CSE, KSIT, Bengaluru. His area of research includes Big Data analytics, Data Mining.

[2]**Dr Naveen N C** received his PhD from SRM University, India. He is currently working as Professor, Department of CSE, JSS Academy of Technical Education, Bengaluru. He has 20 years of work experience in teaching with Research, Administration and Industry. He has published more than 30 papers in national and international conferences and journals. He is guiding PhD students in areas of Big data Analytics, Image processing, Data mining