

MACHINE LEARNING AND CLOUD COMPUTING: A DETAILED STUDY ABOUT DISTRIBUTED AND SOLUTIONS OF SAAS

RADHIKA BAKLIWAL^{*}, MANJU PAYAL^{*}, AISHWARYA MATHUR^{*}, DIVYA MATHUR^{*}

ABSTRACT

For overseeing and examination the substantial measure of information is run of the mill work. So, machine learning calculation is connected to the huge information which prompts the immense test for the ML partitioners. Previous ML libraries was not ready to help well handling of huge datasets, due to which new methodologies were needed. Parallelization is a system which is utilizing modern parallel figuring framework, such as MapReduce, CUDA, or Dryad which are gainful in ubiquity and acknowledgment which brings about growing new ML libraries firstly, than frameworks. We will present the mechanical and scholarly results, like Apache Mahout, GraphWab or Jubatus.

Let's check out how cloud computing model affects the ML. First field direction is towards the popular census tools and collections(R system,Python) is spread out in the cloud. The another line of products is to build up present tools with constituents which allows users to create Hadoop clusters in the cloud which allows platform to run jobs on it. Next, for ML calculation the rundown of libraries of conveyed applications, and for suspicions spread out of complex frameworks for information examination and information mining. Last approach of this review is that ML as programming as-a Service, the a few major information and huge organizations opening their answers in the market.

INTRODUCTION

The quickly development of gathered and accessible information in organizations, science and industry, many technologies are needed for examining this type of facts for obtaining efficient data. Today, for data analyzation no longer controlled on big data and classical databases, but they included huge amount of data such as textual documents and Webpages (Web mining, Text mining), multimedia data, spatial data, relational data (molecules, social networks) Investigation apparatuses permit clients to gather the important examples covered in substantial dimensions of

unstructured and organized information. Breaking down huge datasets enables clients to recognize new income sources, create faithful and profitable client connections, and run your general association all the cost effectively and more efficiently. [4]

Study about machine learning and information revelation joins traditional inquiries of software engineering (well organized calculations, programming frameworks, databases) with parts from artificial learning and estimations desirable customer orchestrated disputes (discernment, instinctive mining.

^{*}MCA Scholar, MDS University Ajmer. **Correspondence E-mail Id:** editor@eurekajournals.com

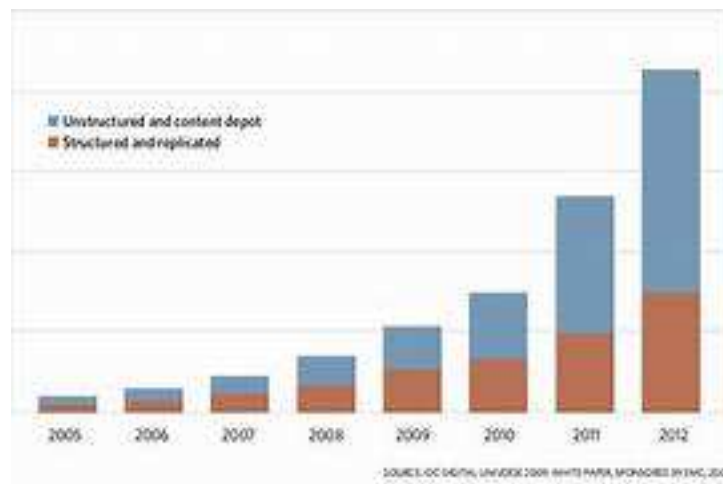


Figure 1. Trends in data growth

In spite of the fact that for over two decades, parallel database items, for example, Teradata, Oracle or Netezza have given intends to understand a similar usage of ML-DM calculations, communicating ML-DM calculations in SQL code is a mind boggling errand and difficult to keep up. In addition, considerable scale foundations of these things are exorbitant and are not an affordable option generally speaking. [8] Another autoist for point of view change from social prototypical to various decisions is the new thought of dataUntil around 5 years prior, maximum information was positioned in nature, comprising of numeric or string information that fit effectively in lines and segments of social records distributed processing environments (MapReduce), or cloud computing.

Now, it merits resonate to a pleasant similitude by Ben Werther prime supporter of Platfora, for huge information handling today:

“In 'mechanical upset' terms, we are in the pre-modern period of artisanship that proceeded large scale manufacturing. It is what might as well be called expecting to connect with a specialist metal forger to produce the forks and spoons for our super table.”

In Machine learning, for overcome the situation

of time exhausting job, many efforts were applied to increase the speed of execution time. Distributed computing model and cloud suppliers designated as an important contrasting options to accelerate machine learning stages. For reducing computation time some popular census tools were applied to the environments like R, Octave, Python-as well as goes in the cloud. [11] There are two principle approaches to merge them with cloud suppliers named as-develop a bunch in the cloud and reboot it with measurement instruments, or conflict measurement conditions with constituents, which give stage to clients to build Hadoop groups in the cloud and performing occupations on those Hadoop groups.

Situations like R, Mapple, Octave and comparative offer low-level foundation for information examination, which could be connected for substantial datasets once utilized by cloud suppliers. Machine Learning goes ahead best of this and encourages the recovery of helpful information available in tremendous information for clients with no/less factual foundation via consequently inducing 'information models' out of information. [2] For helping this requirement, a blast of new companies, some of them in secrecy manner yet, which offers machine learning administrations to the clients, or huge information investigation administrations can

be seen in recent years. These activities can be any PaaS/SaaS stages or items that can be sent on remote conditions. [5]

Evaluating the writing and the market, we can reason that ML-DM comes in numerous flavors. These methodologies are ordered in five unmistakable classes:

- In section 3⇒ Learning conditions from the cloud make a PC bunch in the cloud and bootstrapping it with measurements devices.
- In section 4⇒ Plugins for Machine Learning devices increase insights instruments with modules that enable clients to make a Hadoop group in the cloud and run ML employments on it.
- In section 5⇒ Distributed Machine Learning libraries accumulations of parallelized executions of ML calculations for disseminated situations (Hadoop, Dryad and so forth).
- In section 6⇒ Complex Machine Learning frameworks items that should be introduced on private server farms (or in the cloud) and offers elite information mining and investigation.
- In section 7⇒ Software as a Service suppliers for Machine Learning PaaS/SaaS arrangements that enable customers to get to ML calculations through Web administrations.

The staying of the paper is organized as takes after: succeeding segment presents comparable, late examinations, trailed by five areas, every one of them dedicated to a specific class recognized previously. The paper ends with tentative arrangements and conclusion.

MACHINE LEARNING ENVIRONMENTS FROM THE CLOUD

Suppliers of this class exhibit PC bunches utilizing open cloud suppliers, for example,

Amazon EC2, Rackspace and so on, pre-introduced with insights programming, favored bundles being R framework, Mapple or Octave. [1] These arrangements offer versatile elite assets in the cloud to their clients, who are liberated from the weight of installing and dealing with claim bunches.

CLOUDNUMBERS.COM is consuming supplier of Amazon EC2 to organise the PC groups. For scientific registering these PC groups are reinstated with programming, for example, R framework, Mapple or Octave. Clients got profit from a network interface where clients can make claim workstations; they construct and screen the bunch, transfer the datasets or interface with open databases. Over default highlights from cloud supplier, Cloud numbers offers high security principles by giving secure encryption to information transmission and capacity. Primarily, in cloud a HPC stage, is simple to make and uncomplicated to take up. [6]

CLOUDSTAT is a cloud incorporated advancement condition manufactured in light of R framework, and uncovered its functionalities through 2 kinds of UIs: support-for accomplish customers in R tongue, and applications-sketched out as a opinion and snap shapes based interface for R for customers with no R programming aptitudes. Customers can pick applications from the creating vault from a CloudStat AppStore. [12]

OPANI 4 is award comparative administrations to Cloudnumbers.com, yet moreover causes clients to estimate their bunch as indicated by their requirements: dimension of information and the interval span for preparing this information. They are handled Rackspace's five framework and bolster situations, for example, R framework, Node and Python, packaged with outline, representation, security and variant control packages.

Consequences of information examination forms, named dashboard in Opani, can without much of a stretch be envisioned and shared from work area or cell phones. Methodologies in this class are effective and flexible arrangements, offering clients the likelihood to create complex ML-DM applications kept running on the cloud. Clients are liberated from the weight of provisioning self-conveyed surroundings for scientific processing, while at the same time having the capacity to utilize their most loved conditions. On the opposite side, clients of these devices need broad involvement in programming and solid information of insights. Maybe, because of this restricted gathering of people, the steady suppliers in this class are less than in different classifications, some of them, (for example, CRdata.org) closing down the task just not long after taking off.

DISTRIBUTED MACHINE LEARNING LIBRARIES

This class offers complex libraries working on different dispersed setups (MPI, Dryad, Hadoop). They enable clients to use out-of-the-container calculations, or actualize their self, that are keep running in parallel manner over a group of PCs. These arrangements doesn't coordinate, nor utilize, insights/ arithmetic programming, rather they offer independent bundles of enhanced, state of-the-craftsmanship ML-DM strategies and calculations.

Apache Mahout is an Apache task to make free use of circulated or usually versatile machine learning calculations on the Hadoop stage. [14] It began as an accumulation of autonomous, "without hadoop" parts, e.g. "Taste" community oriented filtering. Its will likely form adaptable machine learning libraries, where versatile has a more extensive importance:

- Ascendable to sensibly expansive datasets.

Mahout's center calculations for alliance, classification and clump construct collective filtering are implemented with respect to best of Apache Hadoop using the guide/diminish worldview. In any case, it does not restrain promises to Hadoop based use: promises that keep working on a single hub or on a non-Hadoop bunch are also welcome. The center libraries are very streamlined to take into consideration great execution additionally for non-circulated calculations.

- Scalable to help different business cases. Mahout is conveyed under a financially inviting Apache Software permit.
- Scalable people group. The objective of Mahout is to fabricate an energetic, responsive, various group to encourage talks on the venture itself as well as on potential utilize cases.

Right now Mahout underpins for the most part four utilize cases:

- Recommendation mining takes clients' conduct and from which attempts to search things clients may like
- Clustering takes example content archives and gatherings them into gatherings of reports related to topics.
- Classification gains from existing arranged archives what reports of a particular class appeared alike and can allot records without label to the (ideally) redress class.
- Frequent itemset mining takes a course of action of thing social occasions (terms in an inquiry session, shopping container content) and identifies, which particular things regularly appear to be as one. [12]

COMPLEX MACHINE LEARNING SYSTEMS

This area display a few answers for business insight and information examination that offer an arrangement of basic highlights: (I) all are

deployable on-start or in-the-cloud groups, (ii) give ironic arrangement of graphical instruments for breaking down, investigate and envision a lot of information, (iii) uncover a fairly restricted arrangement of ML-DM capacities, typically constrained to expectation models and (iv) use Apache Hadoop as preparing motor and additionally stockpiling condition. [4] There are differences on how information is incorporated and handled, bolstered information sources or identified with multifaceted nature of the framework. The maximum identified ones are: Kitenga Analytics, as of late obtained by Dell, is a local Hadoop application that offers graphic ETL, Apache Solr 21-based hunt, characteristic dialect handling, Apache Mahout-based information mining, and propelled representation abilities. It is a noteworthy data condition for refined specialists who require an enthusiastic toolbox of precise gadgets, all from a straightforward touse interface which doesn't need perception of difficult programming or the Apache Hadoop stack itself. Pentaho Business Analytics offers an entire reply for vast data examination, supporting all periods of an examination process, from pre-organizing to cutting edge data examination and observation. [7] It offers (i) a total graphical outline gadget to accelerate data arrangement and representing, (ii) data synchronization from NoSQL and social databases, (iii) circulated implementation on Hadoop stage (iv) moment and intuitive investigation (no ETL (Extract, Transform, Load), no code), and (v) corporate examination level: data exposing, examination, perception and revelatory investigation. Fundamental attributes of Pentaho arrangement include:

- MapReduce-based information handling
- Changes can be made for different Hadoop dispersions, (for instance, Cloudera, Hadapt et cetera.)
- Information can be stacked and prepared into Hadoop Hive, HDFS or HBase

- Backing Pig contents
- Native help for maximum NoSQL databases, for instance, Apache HBase, MongoDB, 10gen et cetera, DataStax, Apache Cassandra
- Allows execution improved information examination, revealing and information joining for logical databases, (for example, Teradata, monetdb, Netezza and so forth.), through profound reconciliation with local SQL lingos and parallel mass information loader
- Integration mind HPCC (High Performance Computing Cluster) from LexisNexis Risk Solutions
- The Import/send out from the PMML (Predictive Modeling Markup Language)
- Pentaho Instaview, a visual application to decrease the time expected to convey information investigation arrangements and to assist amateur clients with getting bits of knowledge of the information, in 3 straightforward improvements: select information source, consequently get ready information for investigation, and picture and investigate assembled models.
- Pentaho Mobile-gives intuitive business examination to corporate clients for iPad applications

Their environment is made out of a few effective frameworks, every one of them its very own intricate task:

Pentaho BI Platform/Server the BI organize is a structure giving focus organizations, for instance, approval, logging, assessing and runs engines; it in like manner has an answer engine that joins each other system (declaring, examination, coordination and data mining); BI Server is considered as a maximum fathomed utilization of the stage, because it works as an electronic report association structure, application mix server and slight workflow motor. [13]

Pentaho Reporting in view of The JFreeReport is a set of open-source devices-The Pentaho Report Designer, Pentaho Reporting SDK, Pentaho Reporting Engine and the normal announcing libraries imparted to the whole Pentaho BI Platform-that enables clients to make social and explanatory reports from an assortment of assets yielding outcomes in different organizations (PDF, HTML Excel and so on.)

The Pentaho Data Integration (Kettle) passes on capable ETL capacities using metadata-driven approach with a natural, rearranged diagram condition, graphical;

Pentaho Analysis Service (Mondrian):= It is an Online Analytical Processing (OLAP) server that help information examination reliably

Pentaho Data Mining (Weka)=: It is a gathering of machine learning calculations for relapse, classification, bunching and affiliation policy;

Platfora passes on in the memory industry understanding with similar information stockroom or ETL necessary. It is a visual interface in perspective of HTML engages business clients to investigate information. Results might be suitably shared between clients. It depends upon Hadoop gathering, that can be displayed either on state start, or on cloud suppliers (S3 and Amazon EMR). It is main spun around BI advantages, for instance, clarified recognition composes (graphs, plots, maps), or cut up undertakings, yet moreover offers a perceptive examination framework. [7]

The Skytree Server is a significant ML and information examination framework that help information from, Hadoop structures, social databases or the flat files and offers connectors to key genuine bundles and machine learning libraries. Machine Learning methods helps to support the Nearest Neighbor, Vector Machine

(SVM), KMeans, Linear Regression, Principal Component Analysis (PCA), 2-point affiliation and KDE (Kernel Density Estimation). For examination front-closes, for example, Web associations or quantifiable and Machine learning libraries (Weka, R), for information depiction. [11] Its arrangement decisions join cloud suppliers, or submitted assembling in context of Linux machines. It in like way underpins clients in reviewing the measure of the social occasion they require by a central condition (Analytics Requirements Index).

Wibidata is a cerebrum boggling course of action in context of open assets programming stack from joining Hadoop, Apache Avro and HBase with select bits. WibiData's ML libraries provide the mechanical gatherings to begin building complex information dealing with pipelines quickly. WibiData moreover gives graphical devices to trade your information from its passed on information store into any social database. [3] With a specific genuine goal to streamline information dealing with utilizing Hadoop, WibiData presents the musings of makers-tally works that resuscitate a fragment in a table, and gatherers-close the hole between WibiData table and key-respect sets organized by Hadoop MapReduce motor. [6]

We realize that we couldn't support the action offer in the field of business knowledge and huge data examination. We hurling to cover the general population who are besides offering machine learning parcels in their uses, different others concentrating just on tremendous information examination, for instance, SiSense, Alteryx, SAP or SAS, being hindered from this survey. Courses of action in this arrangement focus generally business customers, who need to quickly and easily remove bits of learning from their data, being awesome contender for customers with less PC or estimations establishment. [8]

SOFTWARE AS A SERVICE PROVIDERS FOR MACHINE LEARNING



It is segment centers around stage as-an administration, or programming as-a-specialist organizations for machine learning issues. They are offering the administrations primarily by means of RESTful interfaces, and in a few (uncommon) cases the arrangement may likewise be introduced on-commence (Myrrix), differentiating to arrangements from past area that are predominantly deployable frameworks on private server farms. [9] The class of Machine learning issues is the top pick (Google Prediction API, BigML, Eigendog) among these frameworks. We did exclude in this investigation suppliers of SQL over Hadoop arrangements (Example Hadapt, Cloudera Impala Hive) in light of the fact that their primary target isn't ML-DM, rather quick, versatile and adaptable SQL preparing of social information utilizing the circulated design of Hadoop. [3]

BigML: It is a SaaS way to deal with ML. Clients would be able to datasources, make, realize and offer budget models (just choice trees are bolstered), and utilize models to create expectations. Utilize REST API either through automatically or web interface. [14]

BitYota: It is a youthful start-up (2012) SaaS

supplier for Big Data warehousing arrangement. Over information compromise from various basic sources (social, HDFS, NoSQL) it allows clients to run measurements and synopsis questions in SQL, standard R insights and custom capacities written in Perl, JavaScript, or Python on a parallel investigation motor. Results are imagined by coordinating with prevalent BI apparatuses and dashboards.

Precog: It has a more detailed SaaS arrangement made out of Quirrel dialect, Precog database ReportGrid and LabCoat apparatuses. At the center of Precog, we have a unique (no other NoSQL no Hadoop based), schema less, columnar database intended for putting away and examining semi-organized, estimated information, for example, occasions (clients clicking, drawing in, and purchasing), sensor information, movement stream information, actualities, and different sorts of information that don't should be impermanently refreshed. Precog's usefulness is uncovered by REST APIs, however customer libraries are accessible in Python, JavaScript, Ruby, PHP, C# or Java. The LabCoat is a GUI apparatus for creation and administration of Quirrel questions. Quirrel is an a very

meaningful information examination dialect that create it simple to do in-database investigation, insights, and ML over any sort of estimated information. Results are accessible in JSON or CSV groups. ReportGrid is a HTML perception motor that intuitively, or automatically, fabricate reports and outlines. [15]

EigenDog is an administration for versatile prescient displaying, facilitated S3 (for information and models stockpiling) and on Amazon EC2 (for calculation) stages. [5] It assembles choice tree show out of information in Weka's ARFF arrange. Models can be computed in double configuration and incorporated in client uses on account of API, or open-source library gave by merchant.

Metamarkets assert as being Data Science-as-a-Service suppliers, bailing clients to get experiences out of their expansive datasets. [10] They action end-clients the likelihood to implement quick, specially appointed examinations on information, to find new and remarkable irregularities, to spot inclines in information streams, in view of factual models, in an instinctive, intelligent and collaborative

way. They are centered around specialists, less educated on insights and machine learning. Myrrix is an aggregate, continuous, flexible recommender structure manufactured using Apache Mahout. We realize that we couldn't cover all the course of action provider in the field of business knowledge and huge data examination. [16] We attempted to cover the people who are moreover offering ML portions in their applications, various others focusing just on gigantic data examination, for instance, SiSense, Alteryx, SAP or SAS, being blocked from this review. Courses of action in this arrangement focus generally business customers, who need to quickly and easily remove bits of learning from their data, being awesome contender for customers with less PC or estimations establishment.

Predictobot by Prediction Appliance also goes for doing machine getting the hang of showing less requesting. The customer will exchange a spreadsheet of data, answer two or three request, and thereafter download a spreadsheet with the farsighted model. It will pass on judicious showing to anyone with the aptitudes to create a spreadsheet. The industry is still in stealth mode. [8]

TEXT MINING AS SAAS

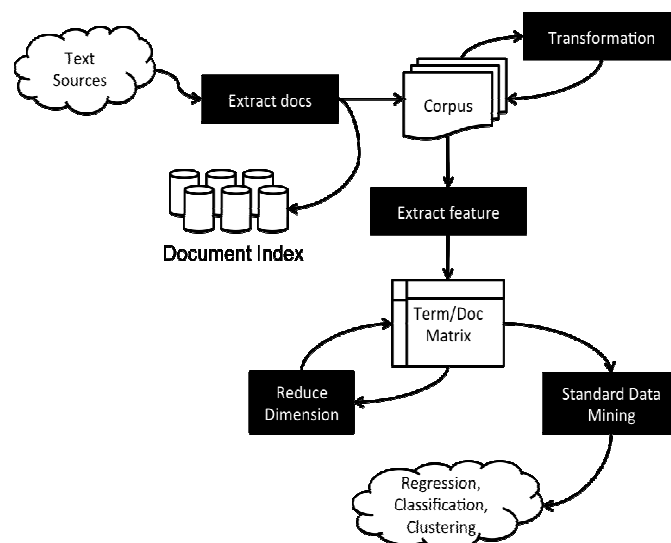


Figure 2.Mechanism of Text Mining

Because of blast of online networking innovations, for example, blog stages (WordPress.com, Blogger and so forth), miniblogging (Twitter), or informal organizations (Facebook, Google+), an expanded intrigue is paid to content mining and regular dialect handling (NLP) arrangements conveyed as administrations to their clients. This is the reason we dedicated a whole subsection to bunch together programming/ stage as-a-benefit answers for content mining. Before surveying accessible arrangements, a short prologue to NLP and content mining is useful. [17] While NLP utilizes etymologically roused methods (content is linguistically parsed utilizing data from a formal sentence structure and a dictionary, and the subsequent data is then translated semantically and used to extricate data) to profoundly examine the report, content mining is later and utilizes systems created in the fields of data recovery, insights, and machine learning. [9] Appearing differently in relation to NLP, content mining's point isn't to comprehend what is "said" in a content, rather to extricate designs crosswise over extensive number of records. Highlights of content mining incorporate extraction of idea/element, content bunching, rundown, or slant examination.

Size and number of records that should be prepared, in addition to ongoing handling oblige add to the advancement of novel, conveyed toolboxes ready to answer requesting clients' needs. Sites administrators are ready to offer content mining highlights to their guests with least venture and decreased support costs. Therefore, an ever increasing number of suppliers are offering content mining administrations through RESTful web administrations, sparing customers from expensive frameworks and organizations. Without going for giving an depth outline of substance mining P(S)aaS supplier, we will indicate a couple of them after this:

AlchemyAPI is a cloud-based substance mining SaaS sort out giving the mainly whole strategy of NLP cutoff points of any substance mining stage, including: named segment taking out, estimation examination, thought checking, relations extraction, producer extraction, page cleaning, tongue affirmation, watchword extraction, references extraction, target mining, and point orchestrate. AlchemyAPI utilizes profound etymological parsing, factual characteristic dialect preparing, and machine figuring out how to investigate your substance, separating semantic meta-information: data about individuals, places, organizations, points, dialects, and the sky is the limit from there. [6] It gives RESTful API limitation, SDKs in all noteworthy programming tongues and answer are encoded in different game plans (JSON, XML, RDF). Relationship with particular facts security require or the regulatory goals is suggested the probability to present the course of action on assert condition. The NathanApp is AI-one's universally useful machine learning PaaS, likewise accessible for organization onpremise as NathanNodeTM. Like Topic- Mapper, it is in a perfect world suited to take in the importance of any human dialect by taking with regards to words, just quicker and with more prominent arrangement extensibility. it is a RESTful API has appropriate JSON and JavaScript.

The TextProcessing is likewise a NLP API that backings originate and lemmatization, slant investigation, labeling and lump extraction, stage extraction and named element acknowledgment. These organizations are suggest open and free (for limited use) by methods for consumer libraries exist in Java, RESTful API endpoints, Ruby, Python, Objective-C and PHP the response are JSON encoded and Python NLTK demos are restricted to achieve a shaky desire to ingest data. For business purposes, customers are offered month to month memberships by means of

Mashape.com. Hurray! Content Analysis Web Service recognizes substances/ ideas, classifications, and connections inside unstructured substance. It positions those identified substances/ ideas by their general significance, settle those if conceivable into Wikipedia pages, and clarifies labels with important meta-information. The organization is open as a YQL table and response is in XML compose. It is uninhibitedly open for non-business usage. [18] This portion showed PaaS game plans having a tendency to, to some degree, machine learning issues. An exceptional sub-region was resolved to content mining issue as a result of its spreading in the scene of ML PaaS scene. We see immense players, for instance, Yahoo! or then again Google, and likewise various new organizations with million dollars fundings. They offer Web designs the probability to easily consolidate in their goals ML information. Simple utilization beat usefulness offered by these administrations, accordingly there are just constrained choices of tweaking calculations behind the administrations. Accordingly, these are great contender for clients with fundamental ML needs, however are not sufficiently flexible for tending to further developed issues. [10]

FUTURE WORK AND CONCLUSIONS

Our fundamental findings are blended beneath:

- a. Current programming measures for bestowing extensive level parallelism, for instance, the Message Passing Interface (MPI) and MR (MapReduce) are real decisions for executing ML-DM checks. More intrigue has been given to MR because of its capacity to oversee tremendous datasets and natural flexibility beside disappointments.
- b. The Machine Learning in appropriated conditions appear in various approaches, offering sensible and price effective diverse other options to standard machine learning

and genuine uses, which are not rotated around spread conditions.

- c. Existing arrangements target either experienced, gifted PC researchers, mathematicians, analysts or tenderfoot clients who are content with no (or couple of) potential outcomes to tune the calculations. Ens-client support and direction is to a great extent missing from existing disseminated ML-DM arrangements.

In the wake of checking on more than 30 different offers available, we imagine that here is a still space for an adaptable, easy to use and convey reply for ML-DM with regard to distributed compute world view, it is focus on the consumers with fewer programming or the factual experience, yet eager to run and modify the progressed scientific ML assignments, for instance, specialists and experts from the fields such as prescription, broadcast communications financial and so forth. To this observe, our feasible arrangements incorporate prototyping like an appropriated structure which is based on the existing circulated MLDM systems, yet improving them with ease of use and simplicity of use highlights.

REFERENCES

- [1]. W. Eckerson-New technologies for Big Data, <http://www.b-eye-network.com/blogs/eckerson/archives/2012/11/new-technologie.php> (2012).
- [2]. B. Werther-Pre-industrial age of big data, June 2012, <http://www.platfora.com/pre-industrialage-of-big-data/>.
- [3]. S. Charrington-Three New Tools Bring Machine Learning Insights to the Masses, February 2012, Read Write Web, <http://www.read-writeweb.com/hack/2012/02/three-new-tools-bring-machine.php>.
- [4]. D. Harris-5 low-profile startups that could change the face of big data, Januray

- 2012, <http://gigaom.com/cloud/5-low-profile-startupsthat-could-change-the-face-of-big-data/>.
- [5]. L. Tierney, A. J. Rossini, Na Li-Snow: A parallel computing framework for the R System, *Int J Parallel Prog* (2009) 37:78-90, DOI 10.1007/s10766008-0077-2.
- [6]. A. Ghoting et al.-SystemML: Declarative machine learning on mapreduce. In *Proceedings of the 2011 IEEE 27th International Conference on Data Engineering, ICDE 11*, pages 231-242, Washington, DC, USA, 2011.
- [7]. Y. Low, J. Gonzalez, A. Kyrola, D. Bickson, C. Guestrin, J. M. Hellerstein-Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud, *Proceedings of the VLDB Endowment*, Vol. 5, No. 8, August 2012, Istanbul, Turkey.
- [8]. S. Hido-Jubatus: Distributed Online Machine Learning Framework for Big Data, *XLDB Asia*, Beijing, 2012 <http://www.slideshare.net/JubatusOfficial/distributed-online-machine-learning-framework-for-big-data>.
- [9]. KD Nuggets Survey 2012, <http://www.kdnuggets.com/software/suites.html>.
- [10]. R. Bekkerman, M. Bilenko and J. Langford (editors)-Scaling up Machine Learning, Cambridge University Press, 2012, summary at http://people.cs.umass.edu/~ronb/scaling_up_machine_learning.htm.
- [11]. S. Owen, R. Anil, T. Dunning, E. Friedman-Mahout in *Action*, Manning Publications, 2011, ISBN 978-1935182689.
- [12]. A. Ghoting, P. Kambadur, E. Pednault, and R. Kannan-NIMBLE: A Toolkit for the Implementation of Parallel Data Mining and Machine Learning Algorithms on MapReduce, *KDD* 11.
- [13]. E. Pednault, E. Yom-Tov, A. Ghoting-IBM Parallel Machine Learning Toolbox, in R. Bekkerman, M. Bilenko and J. Langford (editors)-Scaling up Machine Learning, Cambridge University Press, 2012.
- [14]. M. Isard et al.-Dryad: distributed data-parallel programs from sequential building blocks. In *SIGOPS Operating System Review*, 2007.
- [15]. D. Pop, G. Iuhasz-Survey of Machine Learning Tools and Libraries, Institute e-Austria Timi,soara Technical Report, 2011.
- [16]. Budiu, D. Fetterly, M. Isard, F. McSherry, and Y. Yu-Large-Scale Machine Learning using DryadLINQ, in R. Bekkerman, M. Bilenko and J. Langford (editors)-Scaling up Machine Learning, Cambridge University Press, 2012.
- [17]. Rexer Analytics Survey 2011, <http://www.rexeranalytics.com/Data-Miner-Survey-Results-2011.html>.
- [18]. S. R. Upadhyaya-Parallel approaches to machine learning-A comprehensive survey, *Journal of Parallel and Distributed Computing*, Volume 73, Issue 3, March 2013, Pages 284-292.
- [19]. Sreenivas Sasubilli, Kumar Attangudi Perichiappan Perichappan, P. Srinivas Kumar, Abhishek Kumar, An Approach towards economical hierarchic Search over Encrypted Cloud, pages 125-129; *Annals of Computer Science and Information Systems*, Volume 14. ISSN 2300-5963.