

Data mining techniques for Big data: Purpose, Techniques and Practical Implications

Dr. Neeraj Dahiya¹, Lavi Tyagi², Neeru Sharma³

¹Department of CSE, SRM University, Delhi-NCR, Sonipat, Haryana.

²Department of CSE, Panipat Institute of Engineering & Technology, Panipat, Haryana.

³M. Tech. CSE, Kurukshetra University, Kurukshetra Haryana.

Abstract: The unreasonable development of high dimensional huge information has brought about a more prominent test for information researchers to effectively get important information from these information. Customary Data Mining procedures are not fit to handle large information. Prescient investigation has developed in noticeable quality close by the rise of large information. Association Rule Mining (ARM) is a noteworthy assignment for finding successive examples in Data Mining. It has made extraordinary progress in a plenty of uses, for example, market basket, PC systems, suggestion frameworks, and social insurance. In the previous scarcely any years, developmental calculation based ARM has risen as one of the most well-known examination territories for tending to the high calculation season of conventional ARM. Although various papers have been published, there is no complete examination of existing transformative ARM systems. In this paper, we survey rising examination of developmental computation for ARM. We examine the applications on transformative calculations for various sorts of ARM approaches including mathematical guidelines, fluffy principles, high-utility itemsets, class association rules, and uncommon affiliation rules.

Keywords: Data Mining, Big Data, Distributed computing, Hadoop

I. INTRODUCTION

Mining of Data includes successful information assortment and warehousing just as PC handling. Information digging is utilized for analyzing crude information, including deals numbers, costs, and clients, to grow better showcasing

techniques, improve the presentation or decline the expenses of maintaining the business. Additionally, Data mining serves to find new examples of conduct among purchasers [1].

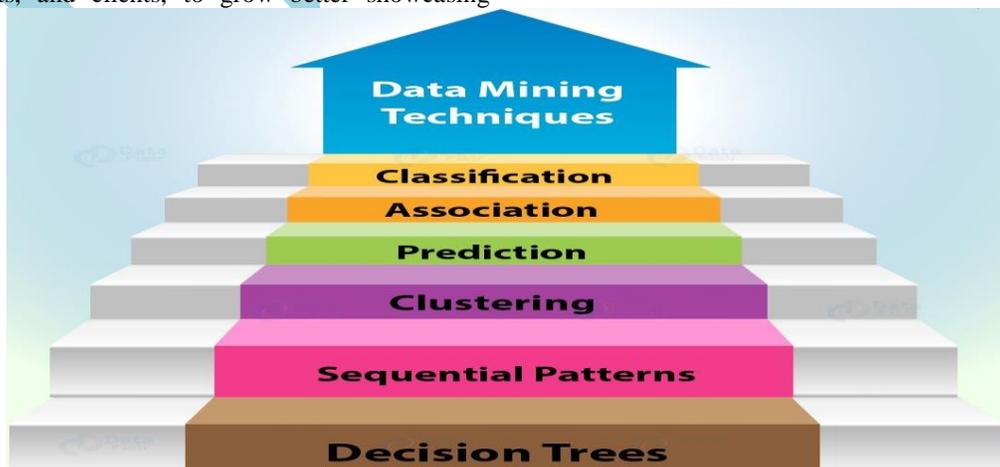


Fig. 1 Data mining techniques

Comprehensively, there are seven primary Data Mining methods.

1. **Statistics:** It is a part of science which identifies with the assortment and portrayal of information. A measurable strategy isn't considered as a Data Mining procedure by numerous investigators. Notwithstanding, it assists with finding the examples and construct prescient models [2].
2. **Clustering:** Grouping is probably the most seasoned strategy utilized in Data Mining. It is the way toward recognizing comparable information that are like one another.
3. **Visualization:** Perception is utilized toward the start of the Data Mining measure [3].
4. **Decision Tree:** This method can be utilized for investigation examination, information preparing and expectation work [4].
5. **Association Rules:** Affiliation Rules help to discover the relationship between at least two things. It assists with knowing the relations between the various factors in information bases. Affiliation rules find the shrouded designs in the informational collections [5].
6. **Neural Networks:** Neural Network is another significant procedure utilized by individuals nowadays. This method is frequently utilized in the beginning phases of the Data Mining innovation. Neural systems are anything but difficult to use as they are robotized to a specific degree and in light of this the client isn't relied upon to have a lot of information about the work or information base [6].
7. **Classification:** This strategy helps in inferring significant data about information and metadata (information about information). Order is firmly identified with the group examination strategy and it utilizes the choice tree or neural system framework [7].

II. DISTINCTION BETWEEN DATA ANALYTICS AND DATA MINING

Information Analytics and Data Mining are two fundamentally the same as controls, both being subsets of Business Intelligence.

- i. Data Mining envelops the connection between quantifiable factors while Data Analytics induces results from quantifiable factors [8].
- ii. Although all types of information examinations are calmly alluded to as "mining of information", there are solid purposes of contrasts between Data Mining and Data Analytics [9].
- iii. Data Mining is utilized to find shrouded designs

among enormous datasets while Data Analytics is utilized to test models and theories on the dataset [10].

- iv. It is the apparatus to improve information for use while Data Analytics helps in creating and chipping away at models for taking business choices. This clarifies why Mining of information depends more on numerical and logical ideas while Data Analytics utilizes business knowledge standards [11].
- v. Data Mining is one of the exercises in Data Analysis. Information Analytics, then again, is a whole range of exercises which deals with the assortment, arrangement, and demonstrating of information for extricating significant experiences or information [12].
- vi. The mining of Data considers are generally founded on organized information. Information Analytics exploration should be possible on both organized, semi-organized and unstructured information [13].
- vii. Data Mining targets making information more usable while Data Analytics helps in demonstrating a theory or taking business choices [14].
- viii. It is generally founded on Mathematical and logical techniques to recognize examples or patterns, Data Analytics utilizes business knowledge and examination models [15].
- ix. This by and large incorporates representation instruments, Data Analytics is constantly joined by perception of results [16].

AI is a subfield of Data Science that centers on structuring calculations that can gain from and make prescient examinations. It includes both Supervised Learning and Unsupervised Learning techniques [17]. Unaided strategies really start off from unlabeled informational indexes, thus, as it were, they are legitimately identified with discovering obscure properties in them (for example groups or rules) [18]. AI can be utilized for Data Mining. Be that as it may, it can utilize different procedures other than or on head of AI [19].

III. BIG DATA

Large information these days have become a trendy expression that is utilized to depict information with enormous volume that can't be prepared utilizing the regular information bases and programming procedures [20]. Information is developing at the exponential speed and it is created from different areas, for example, interpersonal organization (Facebook, Twitter), YouTube, online business (Amazon, Ali Express and so forth.). Different procedures have been created in the writing to deal with and recover better outcomes from these ever-developing information sources [21].

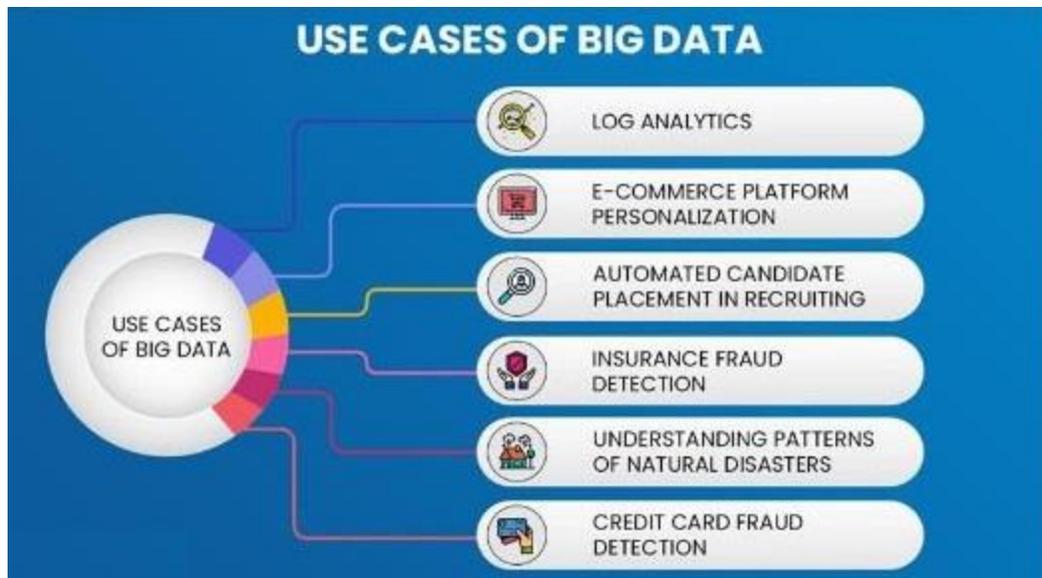


Fig. 2 Use cases of Big Data [22]

The volume is getting greater and greater step by step. Aside from the volume, which has been a significant focal point of the examination network throughout the long term, there are different qualities of Big Data which additionally should be engaged. Laney in 2001 grouped the properties for the Big Data that isolates it from the basic information [23].

As per his hypothesis, the three elements of Big Data are Volume, Variety, and Velocity which can be named as Three V's. Later on, three different measurements, for example, veracity, inconstancy and worth are included and on the whole all these six elements of the Big Data are named as 6 V's. It shows the pictorial portrayal of the 6 V's [24]. All these 6 V's with their own trademark finishes the portrayal of Big Data. Volume means the huge size of the dataset from different areas [25]. The information that is streaming today is of low worth thickness as communicated by Oracle; in any case, viable investigation of information is required for the top notch esteem [26]. This is altogether classified into Value. Assortment is the different configurations of information structure produced from various sources. Changeability is the pace of stream of information at various thickness [27]. Veracity is the vulnerability in the dataset and Velocity is the pace with which information is being produced. Every one of these properties of large information have singular difficulties. From the issue to manage the capacity and the unstructured information, we are likewise confronting bottlenecks in large information speed which continues changing fundamentally and requires various strategies which have high preparing speed [28]. Also, there is issue of information security and pre-preparing of the vague information. An information that is produced from numerous true applications is related with vulnerability,

missing qualities, questionable data, and commotions in the dataset [29]. This is because of the capriciousness of nature, imbalanced condition boundaries, blunder through gadget readings, unstructured design of the information bases and other pointless components [30]. The field of large information is about decade old and throughout the long term it has been seen that fluffy strategies have altogether added to give answers for different huge information issues in a few complex applications [31].

IV. BIG DATA CHALLENGES

In the course of recent decades our period can be depicted as large information time where computerized information is getting progressively significant in numerous areas like medicinal services, science, innovation and society [32]. A lot of information has been caught and produced from different territories, numerous sources, for example, streaming machines, high throughput instruments, sensor systems, versatile application and from each and every field particularly in social insurance, this high information volume speaks to enormous information [33]. Putting away, handling, envisioning and information extraction through this voluminous and differed information types has become a test utilizing lacking best in class advances instruments. One of the most significant innovative difficulties of large information examination is investigating approaches to successfully get important data for various sorts of clients [34]. Presently, the different types of medicinal services information sources are being gathered in both clinical and non-clinical conditions, where the computerized duplicate of a patient's clinical history are the most significant information in social insurance examination [35].

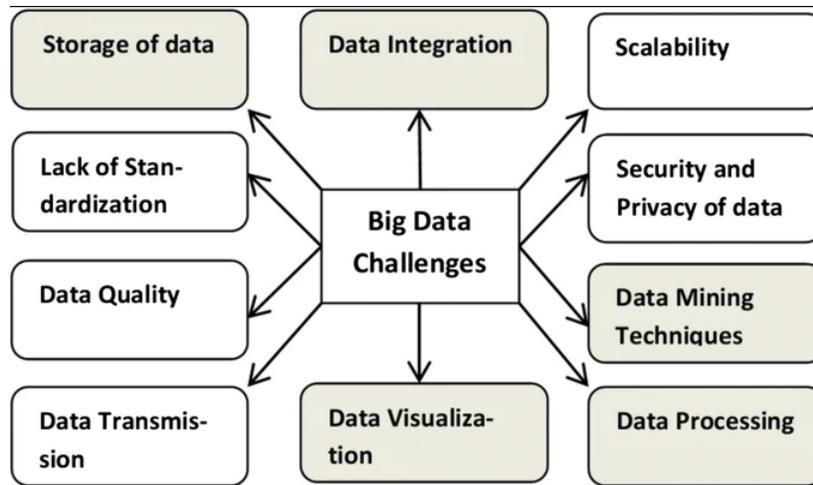


Fig. 3 Big Data Challenges [36]

In this manner, structuring a circulated information framework to manage huge information faces three fundamental difficulties: First, because of the heterogeneous and tremendous volume of information, it is hard to gather information from disseminated areas. Second, stockpiling is the primary issue for heterogeneous and enormous datasets. Enormous information framework needs to store while giving execution ensure [37]. Last test is identified with large information examination, all the more definitely to mining enormous datasets progressively or close to constant that incorporate displaying, perception, forecast, and streamlining [38]. These difficulties require new handling worldview as the current information the board frameworks are not proficient in managing heterogeneous nature of information or the continuous [39]. Notwithstanding, conventional social information base administration frameworks (RDBMS, for example [40], MySQL are predominantly utilized for the board of organized information. These conventional frameworks don't offer any help for unstructured or semi-organized information. From an adaptability viewpoint, when the information size develops, there are numerous standard RDBMS disappointments in scaling for equal equipment the executives and adaptation to internal failure, which isn't appropriate for overseeing developing information [40]. To manage the issues related with huge and heterogeneous information stockpiling, many examination works have been proposed by the exploration network, for example, NoSQL data set administration frameworks [41] which are valuable when working with a colossal amount of information when the information's temperament doesn't need a social model [42].

MapReduce [43] is an equal handling method to deal with enormous information disseminated on an item group; it comprises of the Map and Reduce tasks. It offers a conveyed stockpiling framework by means of its Hadoop Distributed File System (HDFS), it additionally exceptionally deficiency lenient. Hadoop upholds group handling just, it isn't appropriate for continuous stream preparing and in-memory

calculation and it isn't in every case simple to execute the MapReduce worldview for all issues [44]. Contingent upon the volume of the information being handled, the yield can be deferred essentially. Interestingly, stream figuring includes constant information and result of information and it is underlines on the speed of information. Enormous information streaming registering (BDSC) gives high throughput, appropriated messages, ongoing figuring and low-idleness handling [45]. With it's enormously equal handling models, BDSC is a decent decision to increase valuable information from large information which is the key necessity of huge information investigation in social insurance [46].

The quick extension of huge information examines has started to assume a crucial job in the advancement of medicinal services practices and examination. It has given devices to the assortment, the executives, examination and ingestion of a lot of dissimilar, organized and unstructured information created by existing medicinal services frameworks [47]. These days, BDSC assumes a significant job in large information examination to get the shrouded estimation of huge information in social insurance continuously. Anyway due the human services disseminated information sources (the information are originating from the various sources, for example, social data sets, Hadoop, search framework and different examination framework [48]. Applying AI on this large information stream is trying as the conventional AI frameworks are not reasonable to deal with such huge volume or shifted speed. Other issue is identified with the explanatory information preparing. Performing more extravagant expository information preparing includes proficient information mix between frameworks [49]. A large portion of the cutting edge works include AI, however if there should be an occurrence of constant AI applied to streaming enormous information isn't taken care of. Then again the greater part of the medicinal services examination arrangement primarily centered on Hadoop which is a group situated registering [50]. As of late, the quantity of old and residents experiencing ceaseless sicknesses is rising quickly,

impediments of regular wellbeing administrations are turning out to be increasingly significant. Also, the utilization of clinical IoT is expanding for persistent checking so as to perform ongoing activities in the event of crises particularly for coronary illness [51]. In this way, the huge number of sensors produce enormous volume of information. Handling these information and performing continuous activities in basic circumstances is a difficult errand [52].

In view of the difficulties confronting the medicinal services framework we have proposed and built up an answer in social insurance with a continuous wellbeing status expectation use case. These points are put away in a dispersed stockpiling NoSQL for representation and investigation. Productive handling of information in social insurance expands the nature of patient checking [53].

V. Data Mining, AI and information revelation

Information Discovery in Databases (KDD), then again, includes information choice, preprocessing, change, mining, understanding or assessment, and information discovery [54]. Consequently, KDD includes Data Mining as one of the key advances. AI can be viewed as a part of man-made consciousness that utilizes an overall idea of surmising to extricate (learn) the answer for an issue from information samples [55].

Supervised learning

In directed picking up, preparing datasets are marked, and the machine gains from the names to dole out obscure datasets a name upon experience [56]. The outcome is an information dataset planned to a right yield. Accordingly, the term directed learning alludes to oversight by a named preparing dataset to plan the information to an ideal yield [57].

Feature engineering

The high number of highlights in this rundown makes forecast and example distinguishing proof troublesome and at some point less important. Moreover, it makes sense that all qualities from the vector can't be associated with directing transmission [58]. For this situation, include building will be expected to distinguish and choose/diminish (highlight determination, include decrease) this rundown of highlights to the negligible number expected to make applicable expectations [59]. A few methodologies have been actualized for differing information modalities so as to furnish information researcher with area explicit pertinent element distinguishing proof [60].

Qualities and shortcomings of applying Data Mining and AI

DM and ML applications are reforming the field of irresistible sicknesses by adding to early episode discovery, observation, pathogenicity expectation, symptomatic instruments, and then some. Nonetheless, these applications have the two qualities and shortcomings [61]. A specific quality is that frameworks bioscience is delivering a wealth of information that AI and Data Mining techniques can change into novel information [62]. In any case, these omics information come in very heterogenic structures and modalities, making tremendous difficulties for their

utilization, including the "scourge of dimensionality" that relates to huge information. Moreover, issues with missing information, dataset reproducibility, extraordinariness and class lopsidedness, and huge information versatility are among numerous different difficulties [63]. These issues become imperatives and mess up most AI errands when applied to true methodologies, for example, the improvement of clinical arrangements [64]. Thusly, these limitations bring about most models remaining exploration apparatuses for non-clinical and scholarly settings that are helpful just in restricted manners. Regardless, this pattern is evolving and, in this audit, the greater part of the examinations talked about tended to significant certifiable issues [65].

VI. ASSOCIATION RULE MINING.

ARM is the way toward finding ARs in exchange information. It is one of the most critical solo techniques for design acknowledgment [66]

Meaning of ARM

The fundamental ideas of ARM are characterized as follows: An exchange information base (D) comprises of n particular things and m exchanges. $I = \{ I_1, I_2, \dots, I_n \}$ is a lot of n things and $D = (t_1, t_2, \dots, t_m)$ is a limited arrangement of exchanges. An exchange t is a lot of things in a set I , ($t \subseteq I$). An affiliation rule $X \rightarrow Y$ is separated dependent on help and certainty measures [67].

Mathematical/quantitative Association Rule Mining

At that point, BAR mining calculations are applied. At the point when the area of a trait is exceptionally huge or constant, this strategy isn't successful [68]. Discretizing the ceaseless ascribes to various spans is a typical method to address this issue. After discretization, the characteristics are treated as downright traits [69]. For example, a trait x with an incentive somewhere in the range of 20 and 100 can be partitioned into eight spans (20–30, 30–40, ..., 90–100). On the off chance that a worth is 62, the span (60–70) gets 1 and different stretches stay as 0. This test can be tended to utilizing fluffiness. The fluffy set hypothesis [70] is the utilization of normal dialects to speak to the ideas. Due to its effortlessness and similitude to human thinking, fluffy set hypothesis is a famous method in insightful frameworks. Fluffy set hypothesis changes quantitative qualities into semantic terms, in this manner producing etymological or fluffy information [71]

High-utility itemset mining

The cycle of regular itemset mining expects to find an itemset just dependent on the recurrence. This isn't an effective method to discover itemsets related with high benefit on the grounds that itemsets with high utility ordinarily have low help [72], though, visit itemset mining extricates a lot of FIs with low utility. Two measures, backing and utility, are frequently strife with one another [73].

Class Association Rule Mining

Classification and ARM are two broadly utilized strategies in reality. The joining of these models is another pattern in Data Mining [74]. In reality, an arrangement rule is another case of an affiliation rule. ARs have been effectively applied to

order, and the recreation results demonstrated that it could accomplish promising precision. Not at all like an affiliation decide that remembers a few things for the subsequent section, an order rule comprises of the class characteristic just [75].

Rare association rule mining

This is on the grounds that significant data is typically taken cover behind non-visit, irregular, and extraordinary connections. Uncommon ARs (RARs) speak to anomalous or startling practices; hence, they are upheld by just a couple of exchanges. Nonetheless, RARs help to comprehend the new wonders happening in the area [76]. For ex-abundant, in an instructive framework, an educator can check irregular learning issues. An educator finds those understudies who need exceptional help for learning. In medication, RARs assist specialists with recognizing uncommon infections or unfavorable responses of medications [77]. Early examinations tended to the issue of discovering RARs by utilizing successive example mining calculations, a large portion of which were structured dependent on Apriori calculation, which were not viable.

Evolutionary computation for ARM

Because of the idea of high-dimensional spaces, ARM is hard to explain. In this way, conventional heuristic strategies cannot give advanced arrangements, which have brought about expanded fame of non-precise creative improvement approaches known as EC calculations. In the previous, called direction strategies, a calculation begins from an underlying irregular arrangement. At that point, it iteratively produces another arrangement utilizing the bygone one until a predetermined number of emphases are reached.

Evolution-based algorithms

Development based calculations are stochastic hunt strategies dependent on transformative thoughts from characteristic determination and hereditary qualities. These calculations utilize organic administrators, for example, hybrid, change, and characteristic choice. Solid versatility and self-association are the two primary highlights of these calculations. In advancement based calculations, the whole populace is re-set by another age utilizing characteristic administrators, for example, hybrid and transformation. An essential GA comprises of five stages: instatement, assessment, multiplication, hybrid, and change. The regular portrayal of a chromosome is a fixed-length twofold string. The underlying populace is haphazardly created toward the beginning of the GA calculation. Every chromosome is assessed utilizing a wellness work so as to choose people for delivering posterity utilizing a choice component. The administrators of hybrid and change the populace into another one with higher wellness esteems [78]. There are different individual portrayals and wellness capacities.

VII. DIFFICULTIES AND ISSUES

Various issues should be viewed as when creating and applying a developmental ARM calculation. A few issues, for example, huge datasets, property estimations, and boundary settings are identified with both turn of events and

application. Different issues, for example, MFs and arrangement encoding are related with the plan of a calculation. One significant obstruction to utilizing metaheuristic calculations in ARM is that impressive ability and experience are required so as to decide appropriate boundaries, for example, least help and certainty edges just as the number and states of MFs in each quantitative characteristic [79].

Big data

Because of ongoing advances, particularly correspondence innovations, for example, online media, information age capacities have expanded drastically when contrasted with the only remaining century. These days, as everything is put away, a lot of information is being created at a gigantic speed. For instance, when we look for an item utilizing web crawlers, for example, Google, Bing, and Yahoo, we are delivering new information. Large information has four attributes: volume, assortment, speed, and veracity, which are known as the 4V's model. These attributes allude to a gigantic measure of information, different sorts of information, ongoing information, and bad quality information, individually. Albeit developmental ARM calculations can get excellent examples, they battle to unravel large information ARM, in which both the quantity of things and the quantity of exchanges are tremendous. Most customary transformative ARM strategies have concentrated for little scope issues and are not, at this point suitable and can't scale to the size of peta-bytes or zeta-bytes. In this way, it is important to grow new calculations with the capacity to extricate a further extent of information from these data sets. Equal processing is a typical method to improve versatility [80].

Worth sort

Exchange datasets are arranged into two principle gatherings: certain information and questionable information. The extraction of ARs from certain information is direct in light of the fact that the things that happen in an exchange are known with assurance. In this manner, most studies of ARM have concentrated on finding either Boolean/quantitative ARs from certain information or Boolean ARs from probabilistic information. Some datasets in certifiable applications are questionable, in light of the fact that there are various blunders, for example, human mistakes, instrument blunders, and recording mistakes that might be created during the time spent gathering, sending, putting away, and preprocessing information [81]. Information accumulated by sensors are generally unsure because of the dynamicity of nature and their restricted capacity. At times, certain information is changed into dubious information due to the safeguarding of private data. ARM in an unsure information base is more testing than in a specific information base. Traditional ARM calculations, for example, Apriori and FP-Growth can just concentrate ARs from parallel and downright information bases.

Enrollment capacities in fluffy ARM

In fluffy ARM, each numeric characteristic is supplanted by an alternate scope of qualities utilizing the fluffy hypothesis. MFs critically affect the ARM results, where various MFs deduce diverse information [82]. In some developmental ARM calculations, the fluffy sets and their comparing MFs are given by a specialist. Nonetheless, this isn't a proper technique to decide them by specialists for fluffy ARM, since it requires a lot of time and these sets can be changed regarding the measures that the client centers on [83].

Setting least edges

Setting suitable least help and certainty esteems is an unobtrusive errand in ARM [10]. A high edge brings about scarcely any guidelines while low limit causes many (generally repetitive) rules [84]. To address these issues, a few techniques separate ARs without determining estimations of help and certainty. For this situation, just the most fascinating guidelines are found by the intriguing quality measure characterized by the wellness work, rather than extricating an obscure number of intriguing principles. Notwithstanding, this issue prompts various results, for example, creating an immense inquiry space, delivering many invalid chromosomes with decides that don't exist in the dataset, and fundamentally lessening mining productivity. Considering just one fixed help edge for all things isn't reasonable in reality. For example, less expensive things might be set at higher least backings than more costly things. A few analysts have attempted to perform ARM under various help limits for everything. This procedure assists with finding every significant example and even uncommon however significant standards [86].

Arrangement encoding

Arrangement portrayal is unequivocally subject to what kind of Data Mining task should be tended to [12]. There are two notable mappings that encode twofold datasets as far as people: Pittsburgh and Michigan, which were proposed by Holland [15].

VIII. CONCLUSION AND FUTURE WORK

We introduced an arrangement of developmental ARM calculations alongside a concise portrayal of them and a measurable audit of strategies as far as various variables. Applications and basic difficulties were additionally examined. The use of EC calculations to the quantitative ARM has pulled in impressive enthusiasm for ongoing years. Hybridization is another pattern that endeavors to exploit the qualities of various calculations and procedures so as to improve the presentation of metaheuristics or the nature of separated examples. Hybridization isn't limited to the blend of various metaheuristics yet incorporates the mix of metaheuristics and different procedures, for example, fluffy rationale, semantic similitude, and AI calculations including fake neural system and self-sorting out guide. There are a few concerns seeing genuine applications, for example, business, industry, security in PC systems, and suggestion frameworks. These worries are identified with mining valuable and instructive examples regarding recurrence or inconsistency. To upgrade the destinations in these regions, developmental

ARM strategies have been utilized in various manners. Contrasted with different areas, utilization of transformative ARM in industry and business has grabbed the eye of analysts. In certain spaces, for example, instruction, uncommon ARs are removed for distinguishing understudies who need explicit help. Late advances in EC for huge scope worldwide streamlining show that further investigations on this zone are required. In spite of the fact that ARM for huge volumes of information has gained critical ground, no advancement has been made for ARM models of high speed of information. BSO-and GA-based ARM have pulled in more consideration when contrasted with other EC draws near. Different calculations have not thought about the large information issue. Thusly, this region is an examination hole for the Data Mining network. (4) There is an absence of studies assessing the viability of developmental ARM calculations under various viewpoints, which speaks to a magnificent open door for analysts. The impacts of every calculation's boundaries, for example, populace size and backing/certainty edges on the intermingling rate just as getting caught into neighborhood optima can be researched.

References

- [1] Chahar, P., & Dalal, S. (2013). Deadlock resolution techniques: an overview. *Int. J. Sci. Res. Publ*, 3(7), 1–5.
- [2] Dalal, S., & Athavale, D. (2012). Analyzing the supply chain strategy using Case-based Reasoning in *Journal of Supply Chain Management System* ISSN: 2277-1387 Vol. 1 Issue 3.
- [3] Sharma, P., Sharma, D., & Dalal, S. (2015). Preventing Sybil Attack In MANET Using Super Node Using Approach. *International Journal of Recent Research Aspects*, ISSN, 2349–7688.
- [4] Surjeet Dalal, S. (2016). Genetic Algorithm Based Novel Approach for Load Balancing Problem in Cloud Environment. *International Journal of Computer Science and Information Security*, 14(07), 89–93.
- [5] Vishakha, S. (2016). Performance Analysis of Cloud Load Balancing Algorithms. *International Journal of Institutional and Industrial Research*, 1(01), 1–5.
- [6] Shakti Arora, S. (2016). Study of Integrity Based Algorithm in Decentralized Cloud Computing Environment. *International Journal of Institutional & Industrial Research*, 1(01), 15–17.
- [7] Sharma, P., Sharma, D., & Dalal, S. (2014). Preventing Sybil Attack In MANET Using Super Node Using Approach. *International Journal of Recent Research Aspects*, ISSN 2349-7688, 1(1), 30–34.
- [8] Jaglan, V., Dalai, S., & Srinivasan, S. (2011). Enhancing security of agent-oriented techniques programs code using jar files. *International Journal on Computer Science and Engineering*, 3(4), 1627–1632.
- [9] Neeraj Dahiya, S. (2017). Numerical Function

- Optimization: Model, Procedure And Uses. International Journal of Engineering Science and Technology (IJEST), 9(4), 266–270.
- [10] Saini, A., Sharma, K., & Dalal, S. (2014). A survey on outlier detection in WSN. International Journal of Research Aspects of Engineering and Management ISSN, 2348–6627.
- [11] Dalal, S., Jaglan, D., & Sharma, D. (2014). Integrating Multi-case-base-reasoning with Distributed case-based reasoning. International Journal of Advanced Research in IT and Engineering ISSN, 2278–6244.
- [12] Mittal, A., Sharma, K., & Dalal, S. (2014). Applying clustering algorithm in case retrieval phase of the case-based reasoning. International Journal of Research Aspects of Engineering and Management, 1(2), 14–16.
- [13] Dalal, S., Athavale, V., & Jindal, K. (2011). Case retrieval optimization of Case-based reasoning through Knowledge-intensive Similarity measures. Int. J. Comput. Appl, 34(3), 12–18.
- [14] Dalal, S., Tanwar, G., & Alhawat, N. (2013). Designing CBRBDI agent for implementing supply chain system. system, 3(1), 1288–1292.
- [15] Kumar, S., & Dalal, S. (2014). Optimizing Intrusion Detection System using Genetic Algorithm. International Journal of Research Aspects of Engineering and Management ISSN, 2348–6627.
- [16] Sheikh, M., Sharma, K., & Dalal, S. (2014). Efficient method for WiMAX soft handover in VOIP and IPTV. International Journal of Research Aspects of Engineering & Management, 1(2), 5–48.
- [17] Jaglan, V., Dalal, S., & Srinivasan, S. (2011). Improving performance of business intelligence through case based reasoning. International Journal of Engineering Science and Technology, 3(4), 2880–2886.
- [18] Sharma, D., Sharma, K., & Dalal, S. (2014). Optimized load balancing in grid computing using tentative ant colony algorithm. International Journal of Recent Research Aspects, 1(1), 35–39.
- [19] Dalal, S., Jaglan, V., & Sharma, K. (2014). Designing architecture of demand forecasting tool using multi-agent system. International Journal of Advanced Research in Engineering and Applied Sciences, 3(1), 11–20.
- [20] Dalal, S., & Athavale, V. (2012). Analysing supply chain strategy using case-based reasoning. Journal of Supply Chain Management Systems, 1(3), 40.
- [21] Sharma, P., & Dalal, S. (2014). Shortest Path Algorithms Technique for Nearly Acyclic Graphs. International Journal of Recent Research Aspects, 3(3), 36–39.
- [22] Dahiya, N., Dalal, S., & Khatri, S. (2016). Refinement with Image clustering using Self-Organizing Map and Numerical Function Optimization. International Journal of Computer Science and Information Security, 14(11), 909.
- [23] Dahiya, N., Dalal, S., & Tanwar, G. (2016). Refining of image using self-organizing map with clustering. In AIP Conference Proceedings (pp. 020064).
- [24] Mittal, A., Sharma, K., & Dalal, S. (2014). Approach of BPEL in supply chain activities for managing bullwhip effect of SCM system. Int. J. Res. Asp. Eng. Manag, 1(2), 26–30.
- [25] Mittal, A., & Dalal, S. (2014). Implying p-Cure algorithm in case retrieval stage of the case-based reasoning. International Journal of Recent Research Aspects, 3(3), 91–98.
- [26] Jindal, U., & Dalal, S. (2016). Survey on Signature verification and recognition using SIFT and its variant. International Journal of Recent Research Aspects, 3(3), 26–29.
- [27] Dalal, S., Athavale, V., & Jindal, K. (2012). Designing Case-based reasoning applications with Colibri Studio. International Journal of Research in Computer Engineering and Electronics, 1(1), 15–18.
- [28] Dalal, S., Tanwar, G., & Jindal, K. (2012). Agent Oriented Programming In Trading System Automation. International Journal of Research in IT, Management and Engineering, 2(8), 51–59.
- [29] Seth, B., & Dalal, S. (2016). Analysis of cryptographic approaches. International Journal of Recent Research Aspect, 3(1), 21–24.
- [30] Sharma, D., Dalal, S., & Sharma, K. (2014). Evaluating Heuristic based Load Balancing Algorithm through Ant Colony Optimization. environment, 5–9.
- [31] Dalal, S., & Jindal, U. (2016). Performance of integrated signature verification approach. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 3369–3373).
- [32] Jindal, K., Dalal, S., & Sharma, K. (2014). Analyzing spoofing attacks in wireless networks. In 2014 Fourth International Conference on Advanced Computing & Communication Technologies (pp. 398–402).
- [33] Dalal, S., & Sharma, K. (2013). Simulating supply chain activities in multi-agent based supply chain management system with plasma simulator. International journal of Computer Science & Communication, 4(1), 80–85.
- [34] Simi Gupta, D., & Dalal, S. (2014). Efficient broker scheduling in Cloud Computing. International Journal of Recent Research Aspects, 1(2), 74–77.
- [35] Sharma, P., Sharma, K., & Dalal, S. (2014). Preventing Sybil Attack in MANET using Super nodes approach. International Journal of Recent Research Aspects, 1(1), 30–34
- [36] Dalal, S., Jindal, K., & Nirwal, M. (2013). Developing Flexible Decision Support Systems Using Case-Base Reasoning System. International

- Journal of Engineering and Management Research (IJEMR), 3(4), 13–17.
- [37] Arora, S., & Dalal, S. (2017). Adaptive Model For Integrity Verification In Cloud Computing System. International Journal of Advanced Research in Computer Science, 8(1), 233–236.
- [38] Sharma, P., & Dalal, S. (2014). Reviewing MANET Network Security Threats. identity, 25–30.
- [39] Seth, B., & Dalal, S. (2016). Designing Hybrid Security Architecture in Multi Cloud System. International Journal of Control Theory and Applications, 9(41), 767–776.
- [40] Arora, S., & Dalal, S. (2016). Study of Integrity Based Algorithm in Decentralized Cloud Computing Environment. International Journal of Institutional & Industrial Research, 1(1), 15–17.
- [41] Rani, U., & Dalal, S. (2016). Neural Network Applications in Design Process of Decision Support System. International Journal of Recent Research Aspects, 4(2), 40–44.
- [42] Dahiya, N., Dalal, S., & Khatri, S. (2017). An Enhanced Bat Algorithm for Data Clustering Problems.. International Journal of Advanced Research in Computer Science, 8(3).
- [43] Jindal, K., Dalal, S., & Tanwar, G. (2012). Congestion Control Framework in Ad-Hoc Wireless using Neural Networks in QoS. International Journal of Research in Computer Engineering and Electronics, ISSN: 2319-376X, 1(1), 15–18.
- [44] Jindal, K., Dalal, S., & Jaglan, V. (2012). Comparative Study On IEEE 802.11 Wireless Local Area Network Securities.. International Journal of Advanced Research in Computer Science, 3(1).
- [45] Dahiya, N., Dalal, S., Khatri, S., & Kumar, Y. (2016). Cat Swarm Optimization: Applications And Experimental Illustrations To Data Clustering. International Journal of Control Theory and Applications, 9(41), 759–765.
- [46] Arora, S., & Dalal, S. (2016). Novel Approach of Integrity Verification in Dynamic Cloud Environment. International Journal of Computer Science and Information Security, 14(8), 207.
- [47] Dahiya, N., Dalal, S., & Khatri, S. (2017). Data clustering and its Application to numerical function optimization algorithm. International Journal of Advanced Research in Computer Science, 8(1).
- [48] Sharma, S., & Dalal, S. (2014). Recognition and identification schemes for the development of Eigen feature extraction based iris recognition system. International Journal of Recent Research Aspects ISSN: 2349-7688, 1(1), 73-77
- [49] S. Dalal, S. (2017). Integrated approach for performance scrutiny of cloud resource provisioning algorithms. In International Conference of Communication and Computing Systems (pp. 493–498).
- [50] Arora, S., & Dalal, S. (2016). Enhanced Privacy Preserving Access Control in the Cloud. International Journal of Recent Research Aspects, 3(4), 66–70.
- [51] Dahiya, N., Dalal, S., & Khatri, S. (2016). A Review on Numerical function optimization Algorithm and its Applications to Data Clustering & Classification. International Journal of Recent Research Aspects, 3(3), 111–115.
- [52] Rani, U., Dalal, S., & Kumar, J. (2018). Optimizing performance of fuzzy decision support system with multiple parameter dependency for cloud provider evaluation. Int. J. Eng. Technol, 7(1.2), 61–65.
- [53] Surjeet Dalal, V., & Susheel Kumar (2010). Designing of business tool using intelligent agent. In National Conference Advanced Computing & Communication tech ACCT-2010 (pp. 751–754).
- [54] Shakti Arora, S. (2018). Resolving problem of Trust context in Cloud Computing. International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), 5(1), 138–142.
- [55] Dalal, S., & Athavale, V. (2012). Challenging Bullwhip Effect of Supply Chain Through Case Based Multi Agent System: A Review. International Journal of Advanced Research in Computer Science and Software Engineering, 2(12), 267–272.
- [56] Neeraj Dahiya, S. (2016). A Review on Numerical function optimization Algorithm and its Applications to Data Clustering & Classification. International Journal of Recent Research Aspects, 3(3), 115–121.
- [57] Nishant Kumar, N. (2018). Approach of Lion Optimization Algorithm for Efficient Load Balancing in Cloud Computing. International Journal of Management, IT and Engineering (IJMIE), 8(8(1)), 59–72.
- [58] Sameer Nagpal, S. (2018). Comparative Analysis of the Power Consumption Techniques in the Cloud Computing Environment. International Journal of Management, IT & Engineering, 8(8(1)), 32–49.
- [59] Sameer Nagpal, S. (2018). Comparison of Task Scheduling in Cloud Computing Using various Optimization Algorithms. Journal of Computational Information Systems, 14(4), 43–57.
- [60] Nishant Kumar, N. (2018). Approach of Lion Optimization Algorithm and Efficient Load Balancing in Cloud Computing. Journal of Computational Information Systems, 14(4), 32–42.
- [61] Sameer Nagpal, S. (2018). Analysis of LrMu Power Algorithm in the Cloud Computing Environment using CloudSim Toolkit. International Journal of Research in Electronics and Computer Engineering (IJRECE), 6(3), 1175–1177.
- [62] Nishant Kumar, N. (2018). Approach of Lion Optimization Algorithm for Efficient Load Balancing in Cloud Computing. International Journal of Management, IT & Engineering, 8(8(1)), 59–72.

- [63] Le, D.N., Seth, B., & Dalal, S. (2018). A hybrid approach of secret sharing with fragmentation and encryption in cloud environment for securing outsourced medical database: a revolutionary approach. *Journal of Cyber Security and Mobility*, 379–408.
- [64] Sikri, A., Dalal, S., Singh, N., & Dahiya, N. (2018). Data Mining and its Various Concepts. *Kalpa Publications in Engineering*, 2, 95–102.
- [65] Sikri, A., Dalal, S., Singh, N., & Le, D.N. (2019). MAPPING OF E-WALLETS WITH. *Cyber Security in Parallel and Distributed Computing: Concepts, Techniques, Applications and Case Studies*, 245.
- [66] Sikri, A., Dalal, S., Singh, N., & Le, D.N. (2019). Mapping of e-Wallets With Features. *Cyber Security in Parallel and Distributed Computing: Concepts, Techniques, Applications and Case Studies*, 245–261.
- [67] Nagpal, S., Dahiya, N., & Dalal, S. (2018). Comparison of Task Scheduling in Cloud Computing Using various Optimization Algorithms. *Journal of Computational Information Systems* ISSN: 1553-9105, 14(4), 43–57.
- [68] B. Sudha, K. (2019). Early Detection Of Glaucoma Disease In Retinal Fundus Images Using Spatial FCM With Level Set Segmentation. *International Journal of Engineering and Advanced Technology*, 8(5C), 1342–1349.
- [69] Nagpal, S., Dahiya, N., & Dalal, S. (2018). Comparative Analysis of the Power Consumption Techniques in the Cloud Computing Environment. *Journal Homepage: <http://www.ijmra.us>*, 8(8), 1.
- [70] Kumar, N., Dalal, S., & Dahiya, N. (2018). Approach of Lion Optimization Algorithm for Efficient Load Balancing in Cloud Computing. *Journal Homepage: <http://www.ijmra.us>*, 8(8), 1.
- [71] Jindal, U., Dalal, S., & Dahiya, N. (2018). A combine approach of preprocessing in integrated signature verification (ISV). *International Journal of Engineering & Technology*, 7(1.2), 155–159.
- [72] Shakti Arora, S. (2019). DDoS Attacks Simulation in Cloud Computing Environment. *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 414–417.
- [73] Arora, S., & Dalal, S. (2018). Hybrid algorithm designed for handling remote integrity check mechanism over dynamic cloud environment. *International Journal of Engineering & Technology*, 7(2.4), 161–164.
- [74] Shakti Arora, S. (2019). Integrity Verification Mechanisms Adopted in Cloud Environment. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(6S3), 1713–1717.
- [75] Arora, S., & Dalal, S. (2019). Trust Evaluation Factors in Cloud Computing with Open Stack. *Journal of Computational and Theoretical Nanoscience*, 16(12), 5073–5077.
- [76] Arora, S., & Dalal, S. (2019). An Optimized Cloud Architecture for Integrity Verification. *Journal of Computational and Theoretical Nanoscience*, 16(12), 5067–5072.
- [77] Dalal, S., & Kukreja, S. (2016). Genetic Algorithm based Novel approach for Load Balancing problem in Cloud environment. *International Journal of computer science and information security*, 14(7), 88.
- [78] Meghna Hooda, & Shrivankumar Bachu, P. (2020). Artificial Intelligence Technique For Detecting Bone Irregularity Using Fastai. In *International Conference on Industrial Engineering and Operations Management Dubai, UAE* (pp. 2392–2399).
- [79] Seth, B., Dalal, S., Jaglan, V., Le, D.N., Mohan, S., & Srivastava, G. (2020). Integrating encryption techniques for secure data storage in the cloud. *Transactions on Emerging Telecommunications Technologies*, e4108.
- [80] Dalal, S., Agrawal, A., Dahiya, N., & Verma, J. (2020). Software Process Improvement Assessment for Cloud Application Based on Fuzzy Analytical Hierarchy Process Method. In *International Conference on Computational Science and Its Applications* (pp. 989–1001).
- [81] Bijeta Seth, K. (2021). Secure Cloud Data Storage System using Hybrid Paillier-Blowfish Algorithm. *Computers, Materials & Continua*, 67(01), 779–798.
- [82] Dalal, S., & Khalaf, O. (2021). Prediction of occupation stress by implementing convolutional neural network techniques. *Journal of Cases on Information Technology (JCIT)*, 23(3), 27–42.
- [83] Meenakshi Malik, D.N. (2021). Driving Pattern Profiling and Classification Using Deep Learning. *Intelligent Automation & Soft Computing*, 28(3), 887–906.
- [84] Seth, B., Dalal, S., & Dahiya, N. (2021). 4 Practical Implications. *Green Internet of Things for Smart Cities: Concepts, Implications, and Challenges*, 61, CRC Press.
- [85] Dahiya, N., Dalal, S., & Jaglan, V. (2021). 7 Efficient Green Solution. *Green Internet of Things for Smart Cities: Concepts, Implications, and Challenges*, 113, CRC Press.
- [86] Dahiya, N., Dalal, S., & Jaglan, V. (2021). 8 Mobility in Green Management IoT. *Green Internet of Things for Smart Cities: Concepts, Implications, and Challenges*, 125, CRC Press.