

# Data mining techniques in Industrial Engineering: A survey

Dr. Yogesh Kumar Sharma

Associate Professor, Head / Research Coordinator, Department Of Computer Science And Engineering, Shri Jagdishprashad Jhabarmal Tibrewala University, Vidyanagari, Jhunjhunu, Rajasthan-333001

Miss Sunita kumari

M.Tech Student, Department Of Computer Science And Engineering, Shri Jagdishprashad Jhabarmal Tibrewala University, Vidyanagari, Jhunjhunu, Rajasthan-333001

**Abstract**—Information digging offers devices for revelation of connections, examples, and learning in extensive databases. The learning extraction process is computationally unpredictable and in this manner a subset of all information is typically considered for mining. In this paper, various strategies for deterioration of informational collections are talked about. Deterioration upgrades the nature of learning removed from huge databases by rearrangements of the information mining task. The thoughts gave are shown precedents and a modern contextual investigation. For the situation contemplate detailed in this paper, an information mining approach is connected to remove learning from an informational index. The extricated information is utilized for the expectation and aversion of assembling shortcomings in wafers.

**Index Terms**—Data mining, decision making, decomposition, integrated circuit, quality engineering.

## 1. Introduction

There are numerous ways individuals increase new learning. Critical learning is gotten from investigations and viable experience. Understudies are for the most part without down to earth experience yet some can be increased through examining genuine information. In designing educational program, different regular laws, consequences of recreations and tests can be dissected by information mining. By the utilization of information mining innovation on genuine information further learning and experience can be picked up. This encourages instructors and understudies to value certain standards about the examined article. Breaking down genuine information by the methods for information mining is generally utilized in numerous territories of ventures, just as in numerous different zones, including examination and training [1]. The utilization of information mining strategies has been expanding as of late among those information examiners associated with investigating extensive datasets, regular in numerous territories and businesses.



Figure 1

This examination talks about two activities utilizing an information mining grouping calculation to find shrouded leads in datasets. It examines the utilization of grouping in finding the accompanying two considered articles: phone calls and route ways through site pages. We mined phone calls to affirm the conservative purposes behind purchasing a GSM interface for less expensive calling between stationary phones to portable communication. Route designs and the openness of data on website pages were found from information about unknown visits to site pages. This was to pick up learning so as to make better website pages giving data in a simple path with no vital route through a few pages [2].

## 2. Background

Information mining [3] is the unsupervised learning process where:

- new examples and connections can be found in a lot of information,
- new learning is gotten to arrange new information and to discover groups and affiliations,
- new data is found for better basic leadership.

Information mining is a promising innovation that can conquer any hindrance among information and learning. It is a method for learning shrouded leads about the item or procedure depicted by the information. For instruction purposes, it is utilized so as to comprehend and increase some learning of what individuals experience when working in the field.

Information mining does not supplant customary measurable systems [4-9] or useful experience. What are the contrasts between factual techniques and information mining?

- Statistical techniques for acquiring data covered up in information include a client, yet information mining is information driven and does not include the client in the manufacture procedure. Accordingly,

it is helpful for training purposes, where understudies have no or little involvement with information examinations.

- Statistical techniques are manual, while information mining strategies are mechanized. The breaking down and finding of examples and standards covered up in information is finished by the utilization of information mining calculations instead of by the client similar to the case in factual techniques.
- Statistical strategies utilize a top-down way to deal with information examination, while information mining calculations begin with genuine nitty gritty information and they go through a base methodology. In the information mining of little datasets, some incredible examples might be found, however they are just the consequences of specific information tests.

This probability can be perceived in any examination, and this exploratory nature of information mining is vital. It very well may be chosen from training as well as testing different information tests which connections present the genuine connections and which happen in this specific information test.

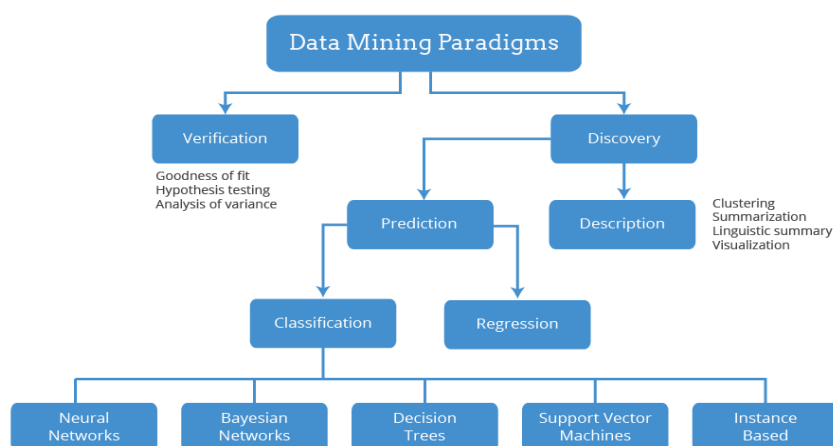


Figure 2

Why use information mining? As of late, various productions have managed the utilization of information mining in various situations, from protection and banking to the synthetic, medicinal, and electrical ventures. The fundamental purposes behind the expanded utilization of information mining strategies are:

- Analyses can be performed rapidly.
- Data mining depends more than conventional factual strategies on the utilization of processing power.
- It can likewise work with uproarious and fragmented information.
- Data mining isn't as reliant as customary examinations on the experience of the information investigators.

Diverse information mining instruments, and fitting calculations dependent on profoundly computerized hunt techniques are accessible. PCs perform a large portion of the information mining work by executing fast calculations.

Every datum mining apparatus must have the accompanying segments:

- Away of perusing information into the information mining apparatus, in the event that it can't utilize the database where the information lives.
- at least one calculations for structure models.
- An interface that underpins client cooperation (generally GUI, in some cases procedural language) and gives an apparatus to the review and assessing of a model.

Information mining strategies are the center of calculations utilized by PCs to give an effective and quick information mining of a lot of information. The best information mining strategies consequently select information and use them in example acknowledgment. They are equipped for working with clamor and fragmented information, and give clear introductions of results to information examiners. There are numerous information digging calculations for arrangement, bunching, and affiliations.

Bunching [10] is an unsupervised learning method for finding characteristic gathering (groups) in the information. It is an extremely prominent and valuable system in information examination and information mining. Bunching calculations make groupings from people having comparative attributes. People can be, for instance, the clients of a business, the understudies of a college, the arrangements of route pages on a web-based interface, or phone calls. Bunching calculations recognize a number people's groupings dependent on their similitudes to one another. Gatherings portray the examined items and are utilized to take a few choices and make exercises, for example,

- Designing distinctive limited time exercises for various client gatherings.
- Different courses can be given to various understudy gatherings based on the found clients' conduct and the found gatherings of calls (with various span and cost) the choice in regards to GSM interface can be made.
- On the outcomes and standards of the route designs, better website pages and web projects can be structured.

Bunching calculations can be separate based.

They depend on a separation metric to gauge the similitudes between information. Groups are ordinarily spoken to either by the mean of the articles doled out to the bunch (k-implies [10-15]) or by one agent object (k-medoid [8]). Information focuses are incorporated into the closest bunch as indicated by the separation metric utilized. Such a bunching calculation is k-implies. A progressive rendition of k-implies assembles the various leveled trees, which can be adjusted or uneven. We built up an information mining device called DM (Fig. 1) with grouping calculation.

We built up a framework based bunching calculation in Cpp. Lattice based strategies initially quantize the bunching space into a limited number of cells, and after that perform grouping on the braced cells. The fundamental preferred standpoint of network based strategies is that their speed just relies upon goals of greeding, however not on the extent of dataset. Groups are recognized utilizing adjusted flood-fill calculation [16]. DM instrument interfaces areas with adequate high thickness into groups. Each group is a most extreme

arrangement of thickness associated cells. All cells having thickness more noteworthy than limit are related into one group. The instrument characterizes edge and lattice. For increasingly exact investigation, the client can characterize them moreover.

The information mining calculation DM is quickly portrayed as pursues:

- perusing information and quantizing the grouping space into a limited number of cells,
- discovering groups by the methods for adjusted floodfill calculation,
- setting up the outcome—groups and number of components of a bunch (named likewise support).

### 3. Historical Development And State-Of-The-Art

Information mining has a long history, with solid roots in insights, computerized reasoning, AI, and database look into (Fayyad, Piatetsky-Shapiro, and Smyth, 1996; Smyth, 2000). Advances in this field were joined by improvement of related programming devices, beginning with centralized server programs for measurable examination in the mid 1950's, and prompting an extensive assortment of independent, customer/server, and online programming as the present administration arrangement. The expression "information mining" was begat in the 1980's and increased developing consideration starting there on (Lovell, 1983).

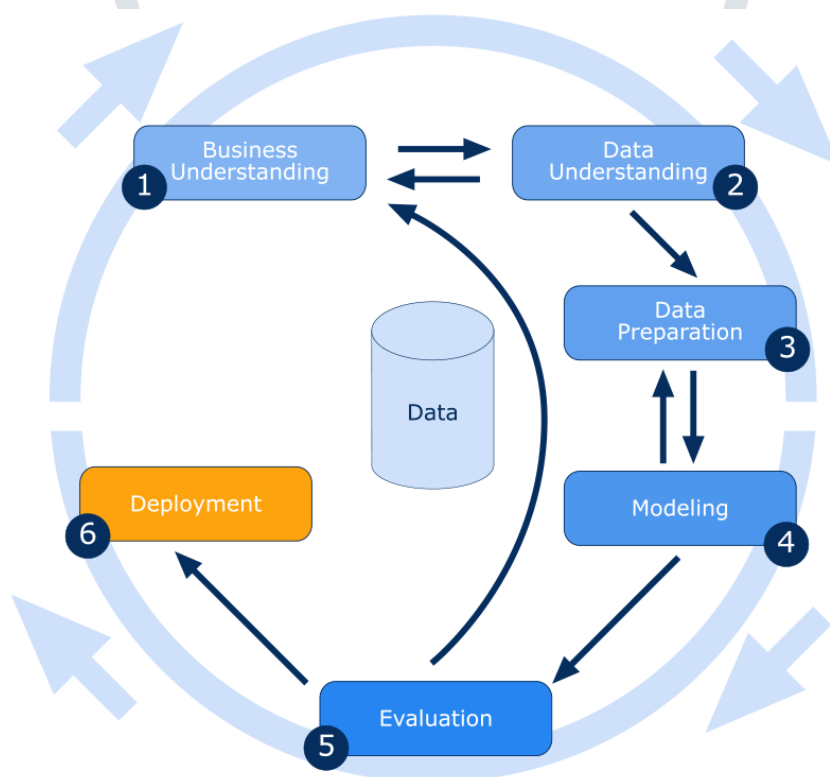


Figure 3

Following the first definition, information mining is a stage in the learning revelation from databases (KDD) process that comprises of applying information examination and disclosure calculations to create a specific identification of examples (or models) over the information (Fayyad et al., 1996). Besides, KDD is characterized as the nontrivial procedure of distinguishing substantial, novel, conceivably helpful, and at last justifiable examples in information. In some cases, the more extensive KDD definition is utilized synonymously for

information mining. This more extensive elucidation is particularly prominent with regards to programming apparatuses, on the grounds that most such devices support the total KDD process, and not only one single step. With ongoing improvements in preparing homogeneous information on circulated frameworks with cluster and stream approaches, a pattern far from the term information mining towards information investigation is conspicuous. Following the meaning of Runkler information examination portrays "the investigation of extensive informational indexes for the help of choices." Furthermore, it is depicted as "interdisciplinary field that has embraced angles from numerous other logical controls (Runkler, 2016)." Today, a substantial number of standard information mining techniques is accessible (Han and Kamber, 2006; Hastie, Tibshirani, and Friedman, 2008). From an authentic point of view, these strategies have diverse roots. One early gathering of techniques was embraced from established insights: the center was transformed from the verification of known theories to the age of new speculations. Models incorporate strategies from Bayesian choice hypothesis, relapse hypothesis, and essential segment investigation. Another gathering of techniques originated from computerized reasoning like guideline based frameworks. The expression "AI" alludes to strategies that can gain from information and incorporates techniques, for example, choice trees, bolster vector machines and counterfeit neural systems. There are a few unique and now and again covering classifications; for instance, fluffy rationale, fake neural systems and transformative calculations are condensed as computational insight (Dalal, 2007).

The run of the mill life cycle of new information mining strategies starts with hypothetical papers dependent on in-house programming models, trailed by open or on-request programming dissemination of effective calculations as research models. At that point, either exceptional business or open-source bundles containing a group of comparative calculations are created, or the calculations are coordinated into existing open-source or business bundles. Numerous organizations have endeavored to advance their very own independent bundles, however just few have achieved remarkable pieces of the overall industry. The existence cycle of certain information mining instruments is strikingly short. Ordinary reasons incorporate inside promoting choices and acquisitions of particular organizations by bigger ones, prompting a renaming and mix of product offerings. The biggest business examples of overcoming adversity came about because of the progression savvy mix of information mining strategies into set up business measurable apparatuses. A short verifiable diagram of vital players can be found in the writer's past article (Mikutand Reischl, 2011). As a rule, measurable instruments and databases with incorporated investigative capacities are mainstream for business application and connected research.

Simultaneously, numerous organizations offering BI items have incorporated information mining arrangements into their database items; one model is Oracle Data Mining (est. 2002). Huge numbers of these items are additionally a result of the obtaining and mix of particular information mining organizations. Reciprocals can likewise be found in the open-source field with databases spent significant time in AI models like MLDB.2 Open-source libraries have additionally turned out to be mainstream since the 1990's. The most unmistakable model is Waikato Environment for Knowledge Analysis (WEKA) (Frank et al., 2005). WEKA began in 1994 as a C++ library, with its first open discharge in 1996. In 1999, it was totally remade as a JAVA bundle; from



that point forward, it has been consistently refreshed. Furthermore, WEKA segments have been incorporated in numerous other open-source devices, for example, Pentaho, RapidMiner and KNIME. An extensive gathering of research models depends on content arranged numerical projects like MATLAB (business) and R (open source). Such scientific projects were not initially centered around information mining, yet contain numerous valuable numerical and perception works that help the execution of information mining calculations. As of late, graphical UIs, for example, those used for R (e.g., Rattle) and MATLAB (e.g., SciXMiner) can be utilized as combination bundles for some, single, open-source calculations [16].



Figure 4

In 2004, Google introduced the programming approach MapReduce, at first utilized for overseeing and examining the expanding measure of sites. Since this methodology was blame tolerant and adaptable to dispersed frameworks, it achieved a high ubiquity. With its combination in the open source venture Apache Hadoop, a model wound up accessible in 2008 and was utilized by Google and Yahoo! (Dignitary and Ghemawat, 2008). For disconnected information handling, it is as yet utilized by numerous organizations like Amazon, IBM, and Facebook. From that point forward, the term huge information has been instituted, alluding not exclusively to the sheer measure of information yet additionally to various information types and quick account and preparing necessities (Wu, Zhu, Wu, and Ding, 2014). This prompts its definition with the three Vs, volume, assortment and speed (Laney, 2001). Expanding this definition with the vulnerability of the information and its business esteem, additionally a definition with five Vs, including veracity and esteem, is conceivable (Ward and Barker, 2013). In 2009, Apache SPARK was introduced to fulfill these new needs and focused on intuitive iterative calculations on groups rather than MapReduce for clump handling. By upgrade of the moderate outcome stockpiling with in-memory calculations and speculation of the MapReduce design with a progressively adaptable coordinated non-cyclic diagram (DAG), SPARK has picked up a high notoriety

(Landset, Khoshgoftaar, Richter, and Dalal, 2013). Particularly the in-memory handling of huge information is a key innovation for quick and responsive mining and can be found in numerous business items like SAP HANA or SAS items (Gan, Lin, Chao, and Zhan, 2017). For AI SPARK offers its very own AI library MLlib however the open source viewpoint additionally permits the improvement of outsider structures like H2O. A review of how enormous information influences AI and traditional information mining was distributed by Japkowicz and Stefanowski (Japkowicz and Stefanowski, 2016). With late patterns like web of things (IoT) and brilliant items or even keen urban areas, information stream investigation and huge information will turn into a significantly progressively important point later on. Consequently, there will be an extraordinary interest for apparatuses permitting to pick up an incentive from this information (Rodríguez-Mazahua et al., 2016). Particularly the investigation of information streams accompanies new difficulties, for example, non-stationary qualities of the information and dynamic conditions including new devices, for example, MOA (Dalal et al., 2013; Gama, 2010). An overview on information stream investigation with an emphasis on troupes can be found in the article of Krawczyk, Minku, Gama, Stefanowski, and Woźniak (2017).

With the expanding preparing intensity of conveyed frameworks, additionally inquire about in explorative scan for information handling pipelines is conceivable. Condensed by the term AutoML (Automatic Machine Learning), there are first endeavors to completely robotize the procedure of calculation choice and parameterization, similar to the device AutoWEKA (Kotthoff, Thornton, Hoos, Hutter, and Leyton-Brown, 2016). Moreover, the programmed arrangement of single techniques with high multifaceted nature like convolutional neural systems is a subject of research (Feurer et al., 2015; Suganuma, Shirakawa, and Nagao, 2017; Wever, van Rooijen, and Hamann, 2017).

The expanding measure of accessible information influences the devices as well as offers ascend to diagnostic techniques like profound learning (DL) (LeCun, Bengio, and Hinton, 2015). Particularly in benchmarks for picture mining these progressions ended up perceptible. In 2012, the primary group to utilize DL was the main group to accomplish a mistake rate of under 25% in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015). Starting here on, most of members concentrated on DL methodologies and accomplished a blunder rate of under 5% in 2017. These advancements can likewise be seen in medicinal picture examination, where commitments of DL have been connected with progress and are required to assume a considerably greater job later on (Litjens et al., 2017).

As DL requires expansive datasets for preparing, but on the other hand is computationally costly, systems utilizing illustrations handling units (GPU) wound up prevalent. The main well known structure in this field was Theano, acquainting the idea with express expansive models as numerical articulations and utilizing calculation charts for execution and memory advancement (Theano Improvement Team, 2016). Today, the greater part of the product utilize a Cuda/C++ based backend and accompany APIs for Industrial applications [17-25].



## 4. Techniques

Since information mining presents many testing research issues, direct utilizations of strategies and procedures created in related examinations in AI, measurements, and database frameworks can't take care of these issues. It is important to perform committed examinations to create new information mining strategies or create coordinated systems for proficient and viable information mining. In this sense, information mining itself has framed an autonomous new field

### 4.1 Classifying Data Mining

There have been numerous advances on explores and improvements of information mining, and numerous information mining methods and frameworks have as of late been created. Distinctive order plans can be utilized to classify information mining strategies and frameworks dependent on the sorts of databases to be contemplated, the sorts of learning to be found, and the sorts of methods to be used, as appeared as follows.

#### **What sorts of databases to take a shot at.**

An information mining framework can be grouped by the sorts of databases on which the information mining is performed. For instance, a framework is a social information excavator in the event that it finds learning from social information, or an item arranged one on the off chance that it mines learning from article situated databases. As a rule, an information digger can be grouped by its mining of learning from the accompanying various types of databases: social databases, exchange databases, objectoriented databases, deductive databases, spatial databases, worldly databases, mixed media databases, heterogeneous databases, dynamic databases, inheritance databases, and the Internet data base.

#### **What sort of learning to be mined.**

A few ordinary sorts of learning can be found by information diggers, including affiliation rules, trademark rules, order rules, discriminant rules, bunching, development, and deviation investigation, which will be examined in detail in the following subsection. Besides, information diggers can likewise be sorted by the reflection dimension of its found learning which might be arranged into summed up learning, crude dimension information, and multiplelevel learning. An adaptable information mining framework may find learning at numerous reflection levels.

#### **What sort of strategies to be used.**

Information excavators can likewise be classified by the fundamental information mining strategies. For instance, it tends to be sorted by the determined strategy into self-governing information digger, information driven excavator, question driven digger, and intuitive information excavator. It can likewise be classified by its hidden information mining approach into speculation based mining, design based mining, mining dependent on insights or numerical hypotheses, and incorporated methodologies, and so on.

Among a wide range of characterization conspires, this review pursues fundamentally one grouping plan: the sorts of information to be mined in light of the fact that such an arrangement shows a reasonable picture on various information mining prerequisites and methods. Strategies for mining various types of information, including affiliation rules, portrayal, grouping, bunching, and so on., are analyzed top to bottom. For mining a

specific sort of information, distinctive methodologies, for example, AI approach, measurable methodology, and extensive database-situated methodology, are analyzed, with an accentuation on the database issues, for example, productivity and versatility.

#### 4.2 Mining Different Kinds of Knowledge from Databases

Information mining is an application-subordinate issue and diverse applications may require distinctive mining strategies to adapt to. As a rule, the sorts of learning which can be found in a database are ordered as pursues. Mining affiliation runs in value-based or social databases has as of late pulled in a ton of consideration in database networks [4], [7], [39], [57], [66], [73], [78]. The assignment is to determine a lot of solid affiliation manages as "A<sub>1</sub> A ... A<sub>n</sub> A, \* E<sub>1</sub> A ... A<sub>B</sub>," where A<sub>i</sub> (for  $i \in \{1, \dots, m\}$ ) furthermore, B<sub>j</sub> (for  $j \in \{1, \dots, n\}$ ) are sets of trait esteems, from the significant informational indexes in a database. For instance, one may discover, from a huge arrangement of exchange information, such an affiliation rule as though a client gets (one brand of) milk, he/she generally purchases (another brand of) bread in a similar exchange.

Since mining affiliation standards may require to over and over sweep through an expansive exchange database to discover diverse affiliation designs, the measure of handling could be colossal, and execution improvement is a basic concern. Productive calculations for mining affiliation standards and a few techniques for further execution improvements will be inspected in Section 3 [17].

The most prominently utilized information mining and information investigation devices related with database framework items are information speculation and synopsis instruments, which convey a few elective names, for example, on-line expository preparing (OLAP), different dimensional databases, information solid shapes, information deliberation, speculation, rundown, portrayal, and so on. Information speculation and synopsis displays the general qualities or an abridged abnormal state see over a lot of client determined information in a database. For instance, the general qualities of the specialized staffs in an organization can be portrayed as a lot of trademark rules or a lot of summed up rundown tables. Besides, usually attractive to show summed up perspectives about the information at various reflection levels. A diagram on staggered information speculation, rundown, and portrayal is displayed in Section 4.

Another vital utilization of information mining is the capacity to perform order in a gigantic measure of information [18-20].

This is alluded to as mining characterization rules. Information grouping is to order a lot of information dependent on their qualities in specific characteristics. For instance, it is alluring for a vehicle vendor to characterize its clients as per their inclination for autos so the business people will realize whom to approach, and inventories of new models can be sent straightforwardly to those clients with recognized highlights to boost the business opportunity. A few investigations in order guidelines will be checked on in Section 5.

In Section 6, we examine the procedures on information bunching. Fundamentally, information bunching is to amass a lot of information (without a predefined class property), in light of the calculated grouping guideline: boosting the intraclass closeness and limiting the interclass comparability. For instance, a lot of ware items can be first bunched into a lot of classes and afterward a lot of standards can be inferred dependent on such an order.

Such bunching may encourage scientific categorization arrangement, which implies the association of perceptions into a progressive system of classes that bunch comparative occasions together. Worldly or spatial-transient information establishes a substantial part of information put away in PCs [91, [SO].



Figure 5

Instances of this sort of database include: budgetary database for stock value record, therapeutic databases, and sight and sound databases, to give some examples. Hunting down comparable examples in a transient or spatial-fleeting database is fundamental in numerous information mining activities [1], [3], [56] so as to find and anticipate the hazard, causality, and pattern related with a particular example. Ordinary inquiries for this kind of database incorporate distinguishing organizations with comparative development designs, items with comparable selling designs, stocks with comparable value development, pictures with comparative climate designs, topographical highlights, natural contaminations, or astrophysical examples. These inquiries constantly require comparability coordinates rather than careful matches. The methodology of example based closeness seek is looked into in Section 7.

In a conveyed data giving condition, records or items are typically connected together to encourage intuitive access. Understanding client get to designs in such situations won't just help improving the framework structure yet in addition have the capacity to prompt better showcasing choices. Catching client get to designs in such conditions is alluded to as mining way traversal designs. Notice, in any case, that since clients are going along the data giving administrations to hunt to the ideal data, a few articles are visited due to their areas as opposed to their substance, appearing very contrast between the traversal design issue and others which are for the most part dependent on client exchanges. The capacity of mining way traversal designs is talked about in Section 8 [26-30].

Notwithstanding the issues considered above, there are positively numerous different angles on information mining that merit examining. Usually important to utilize an information mining inquiry language or graphical UI to determine the fascinating subset of information, the significant arrangement of properties, and the sorts of principles to be found. In addition, usually important to perform intelligent information mining to look at, change, and control middle information mining results, center at various idea levels, or test various types of limits. Visual portrayal of information and learning may encourage intuitive learning mining in databases [31-32].

## 5. Conclusion

As the PC business experiences an amazing period of acquainting new advances and arrangements with the data preparing commercial center, new headings in information mining and choice help are being empowered. One huge empowering influence is the Internet. As associations depend on the Internet for data trade, information mining arrangements are being made that will fit consistently in this new medium. One application territory is content mining, which endeavors to apply the conventional information mining calculations however with regards to nontabular, unstructured information, for example, record accumulations. Bunching and prescient displaying calculations have been effectively connected to issues, for example, report ordering and point recognizable proof in the exploration setting.

Applications dependent on these thoughts are currently developing in the commercial center. Another zone of new applications is the examination and investigation of Internet traffic. Similarly as deals and bank information could be mined to enable the retail to store or bank improve its items and promoting, Internet traffic on a Web webpage can be investigated to more readily comprehend where the genuine interest is, the thing that pages are being taken a gander at altogether, etc. Specialist co-ops can utilize this data to all the more likely arrange their Web pages. At last, there is the likelihood of mining over the Internet. Sellers would offer their frameworks and calculations as Internet servers, and customers would utilize these servers by means of electronic charge based access to mine their information. Early models are being benchmarked in the lab, and items will before long enter the commercial center.

Research in the hidden calculations is a long way from done. Many open issues remain. On the off chance that we are to utilize adaptability, exactness, heartiness, and interpretability as the criteria by which to pass judgment on information mining calculations, at that point no current calculations exceed expectations at the same time in all criteria. This keeps on being the Holy Grail for information mining calculation explore. Will existing procedures be adjusted, or new calculations structured, that are versatile (so the measure of information doesn't represent an issue), vigorous (so they function admirably in a wide assortment of areas), exact (so data separated from the information keeps on holding up outside and past the quick information), and interpretable (giving understanding and incentive to clients)? Moreover, explore proceeds in expanding and adjusting information mining calculations so they can work on a more extravagant gathering of information types. Information is never again simply numerical or discrete. It might be unstructured content, or video, or sound, and the gathering of these more up to date information types is significantly developing. Creating digging methods for separating helpful learning from this new differing wellspring of information will keep investigate murmuring into what's to come.

## References

- [1]. M. James, Classification Algorithms, John Wiley & Sons, New York, 1985.
- [2]. H. Scheffe, The Analysis of Variance, John Wiley & Sons, New York, 1959.

- [3]. D. Michie, D. Spiegelhalter, and C. Taylor, Machine Learning, Neural and Statistical Classification, Ellis Horwood, New York, 1994.
- [4]. B. Ripley, Pattern Recognition and Neural Networks, Cambridge Univ. Press, Cambridge, UK, 1996.
- [5]. S. Weiss and C.A. Kulikowski, Computer Systems that Learn, Morgan Kaufmann, San Francisco, 1991.
- [6]. A.K. Jain and R.C. Dubes, Algorithms for Clustering Data, Prentice Hall, Englewood Cliffs, N.J., 1988.
- [7]. L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, New York, 1990.
- [8]. R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM SIGMOD Conf. Management of Data, ACM, New York, 1993, pp. 207-216.
- [9]. U. Fayyad et al., eds., Advances in Knowledge Discovery and Data Mining, AAAI Press/MIT Press, Cambridge, Mass., 1995.
- [10]. P. Winston, Artificial Intelligence, 2nd ed. Reading, MA, USA: Addison-Wesley, 1984.
- [11]. N. Caetano, P. Cortez, and R. M. Laureano, "Using data mining for prediction of hospital length of stay: An application of the CRISP-DM methodology," in Enterprise Information Systems. Cham, Switzerland: Springer, 2014, pp. 149166.
- [12]. Y. Chen, Y. Li, X.-Q. Cheng, and L. Guo, "Survey and taxonomy of feature selection algorithms in intrusion detection system," in Information Security and Cryptology. Berlin, Germany: Springer, 2006, pp. 153167.
- [13]. O. Banos et al., "An innovative platform for person-centric health and wellness support," in Proc. Int. Conf. Bioinformat. Biomed. Eng., 2015, pp. 131140.
- [14]. W. A. Khan et al., "Mining minds: Journey of evolutionary platform for ubiquitous wellness," in Proc. 12th Int. Conf. Ubiquitous Healthcare (u-Healthcare), 2015, pp. 13.
- [15]. R. Stoean and F. Gorunescu, "A survey on feature ranking by means of evolutionary computation," Ann. Univ. Craiova-Math. Comput. Sci. Ser., vol. 40, no. 1, pp. 100105, 2013.
- [16]. S. Whiteson, P. Stone, K. O. Stanley, R. Miikkulainen, and N. Kohl, "Automatic feature selection in neuroevolution," in Proc. 7th Annu. Conf. Genetic Evol. Comput., 2005, pp. 12251232.
- [17]. S. Doraisamy, S. Golzari, N. Mohd, M. N. Sulaiman, and N. I. Udzir, "A study on feature selection and classification techniques for automatic genre classification of traditional malay music," in Proc. ISMIR, 2008, pp. 331336.
- [18]. Surjeet Dalal & Pooja Chahar, Deadlock Resolution Techniques: An Overview in International Journal of Scientific and Research Publications (IJSRP), ISSN 2250-3153, Volume 3, Issue 7, July 2013, pp. 1-6.
- [19]. Surjeet Dalal, Dr. Vijay athavale, Dr. RC Chhipa, Designing supply chain strategy for water utilization using Case-based Reasoning in SGVU International Journal of Engineering and Technology ISSN: 2277-6915 Vol. 1 Issue 2, pp. 23-28, July 2013.



- [20]. Surjeet Dalal, Keshav Jindal, Monika Nirwal, Developing Flexible Decision Support Systems Using Case-Base Reasoning System in International Journal of Engineering and Management Research, Vol.-3, Issue-4, August 2013, ISSN No.: 2250-0758, Pages: 13-17.
- [21]. Surjeet Dalal, Keshav Jindal and Dr. S Srinivasan, Agent-based Decision Support System using Case-based Reasoning in International Journal of Advanced Research in Computer Science and Software Engineering ISSN: 2277-128X, Vol. 3, Issue 11, November 2013, pp. 879-893.
- [22]. Surjeet Dalal, Dr. Gundeep Tanwar, Dr. Kamal Kumar Sharma, BPEL based scheduling in multi-agent system for business applications in International Journal of IT & Knowledge Management (ISSN 0973-4414), Volume 7, Number 1, December 2013 pp. 103-107.
- [23]. A. Sharma and S. Dey, "Performance investigation of feature selection methods and sentiment lexicons for sentiment analysis," IJCA Special Issue Adv. Comput. Commun. Technol. HPC Appl., vol. 3, pp. 1520, Jun. 2012.
- [24]. S. I. Ali and W. Shahzad, "A feature subset selection method based on symmetric uncertainty and ant colony optimization," in Proc. Int. Conf. Emerg. Technol. (ICET), 2012, pp. 16.
- [25]. A. Ahmad and L. Dey, "A feature selection technique for classi-factory analysis," Pattern Recognit. Lett., vol. 26, no. 1, pp. 4356, 2005.
- [26]. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res., vol. 3, pp. 11571182, Jan. 2003.
- [27]. M. Attik, "Using ensemble feature selection approach in selecting subset with relevant features," in Proc. Int. Symp. Neural Netw., 2006, pp. 13591366
- [28]. Dr. Yogesh Kumar Sharma (2018), "Enhanced Technique for LSB Based Security in Digital Color Images Using Visual Cryptography", "Journal of Computational Information Systems", ISSN: 1553-9105, Vol. 14, Issue 6, Pp. 81-88.
- [29]. Dr. Yogesh Kumar Sharma (2018), "Designing enhanced Security Architecture for 5G Networks", International Journal of Management, IT & Engineering, ISSN: 2249-0558, Vol. 8, Issue 8(1), Pp. 73-83.
- [30]. Dr. Yogesh Kumar Sharma and P. C. Harish (2018), "Critical Study of Software Models Used Cloud Application Development", International Journal of Engineering & Technology, E-ISSN: 2227-524X, Vol. 7, Issue 3.29, Pp. 514-518.
- [31]. Dr. Yogesh Kumar Sharma and Dr. Surender (2013), "A Comparative Performance Study of Bluetooth and Zigbee Protocols", "Research Reformer – International Referred Online Research Journal", ISSN-2319-6904, Issue No. XI, Pp. 3-17
- [32]. Dr. Yogesh Kumar Sharma (2018), "Framework for Privacy Preserving Classification in Data Mining", "Journal of Emerging Technologies and Innovative Research", ISSN: 2349-5162, Vol. 5, Issue 9, Pp. 178-183.