Data mining techniques and applications — A decade review

Conference Paper · September 2017

DOI: 10.23919/IConAC.2017.8082090

3 authors:

Hussain Ahmad
COMSATS University Islamabad
2 PUBLICATIONS **18** CITATIONS
SEE PROFILE

Zahid Anwar
Bahria University
3 PUBLICATIONS **21** CITATIONS
SEE PROFILE

Munam Ali Shah
COMSATS University Islamabad
174 PUBLICATIONS **1,053** CITATIONS
SEE PROFILE

Data Mining Techniques and Applications – A Decade Review

Hussain Ahmad Madni
Department of Computer Science,
COMSATS Institute of Information Technology, Islamabad, Pakistan
hamadnig@gmail.com

Zahid Anwar
Department of Computer Science,
Bahria University Islamabad, Pakistan
ranazahid0342@gmail.com

Munam Ali Shah
Department of Computer Science,
COMSATS Institute of Information Technology, Islamabad, Pakistan
mshah@comsats.edu.pk

Abstract: — Data mining is also known as Knowledge Discovery in Database (KDD). It is also defined as the process which includes extracting the interesting, interpretable and useful information from the raw data. There are different sources that generate raw data in very large amount. This is the main reason the applications of data mining are increasing rapidly. This paper reviews data mining techniques and its applications such as educational data mining (EDM), finance, commerce, life sciences and medical etc. We group existing approaches to determine how the data mining can be used in different fields. Our categorization specifically focuses on the research that has been published over the period 2007-2017. With this categorization, we present an easy and concise view of different models adapted in the data mining.

Keywords— Educational Data Mining (EDM), Knowledge Discovery in Database (KDD), Learning Management System (LMS), Social Network Analysis (SNA).

I. INTRODUCTION

Data mining techniques (DMT) are used to transform raw data to useful information or knowledge. Data itself is nothing, but to process it, is very useful and interesting [1]. There are many advance technologies that use data as useful information intelligently. For example, Knowledge Discovery in Database (KDD) is the process of required output extraction in different formats from raw data. KDD is also defined as the process to view useful patterns in data [2]. A generic and most common diagram of data mining or KDD is shown in Fig.1.

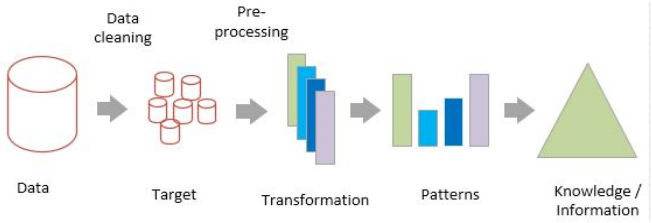

Figure 1: Knowledge discovery from raw data

EDM is the major application of DMT. It is used to produce such datasets which automate the decision of students and administrators. It is mentioned as an emerging program in education which explores many types of data produced by the educational institutions. It analyzes the data generated through the educational system, and improves the learning and educational effects [3]. It belongs to the literature that is related to data mining, visualization, machine learning and computation. Moreover, methods used in machine learning are Naïve Bayes, Neural Networks, K-Nearest Neighbor, Decision Trees and many others [4]. Other fields are also merged with data mining. For example, there are many proposed approaches that are the combination of data mining and semantic web [5]. Similarly, data mining techniques along with machine learning are used in many applications [6].

Large volume of data is produced from web based e-learning which is common today. This huge amount of data is generated by web servers that can be different obtained from multiple web servers. These data can contain user information, location, login history and activities. There are two main sources of data production in EDM that are traditional classrooms and distance education. In case of classrooms, teachers and students are present physically. It involves higher, elementary, primary, private and public education. Educators observe students' behavior by attendance, course information, exams, curricular activities and planning. Educational data mining helps each individual associated with institution. For example, institutions need to know which student can be enrolled in particular course. Administration needs information like class enrollment size and admission requirements. Student needs to understand how to choose or select courses based on prediction which course is the best. Instructors need to know which teaching experiences are the best and most contributive to the class. Educational systems generate data travelling online through web. Data contain student information, course information, administrative information and academic information. Data mining techniques (classification, clustering, text mining, pattern matching etc.) are applied on the data obtained from web. Not only in education, data mining is used by all departments like administration, accounting, Human Resource (HR) and many more. Table 1, shows different tools used in EDM along with their task.

TABLE 1: Data Mining Tools

Tool Name	Mining Task
Mining tool	Association and patterns
MultiStar	Association and classification
Data Analysis Center	Association and classification
EPRules	Association
KAON	Text mining and clustering
TADA-ED	Classification and association
O3R	Sequential patterns
Synergo/ColAt	Statistics and visualization
GISMO/CourseVis	Visualization
Listen tool	Visualization
TAFPA	Classification
iPDF-Analyzer	Text Mining

II. RESEARCH WORK IN DATA MINING

As data mining has become most popular and its use has become most common. It makes automated systems by applying different data mining techniques to data flow. Many algorithms are applied in data mining techniques to solve real life problems. There are many advantages of data mining like it is helpful in banking, finance, accounting, retail, marketing, manufacturing, governments and many more [7]. In the same way, it also has many disadvantages as there are security issues, privacy issues, misuse of information, use of inaccurate information, risk of data loss etc. With the passage of time, data mining is growing and has been improved. There are many journals and articles written about it. Data mining can be used in different perspective with respect to dataset given to solve a specific real time issue [8].

In case of distance education, different techniques are applied to grant access to the students who are far from space and time of lectures in traditional class rooms. Distance education involves internet education, web-based education, multimedia education and videotape education. Web-based education is common among all of them. This type of education creates the history of users' accesses in web logs [9].

It is compulsory to convert the data into particular format to use in a suitable data mining algorithm [10]. Some important processes used to format the data before implementation of data mining algorithms, are given in Table 2.

TABLE 2: Data formatting processes

Process name	Objective
Data Cleaning	Irrelevant data is removed from the raw data. Only useful data is left that is needed for the specific mining algorithm
User identification	Referring the specific page to the user associated
Transaction identification	It makes smaller units of sessions as per transaction
Data transformation	It is creation of new attributes from the existing data
Data integration	Data is integrated and synchronized
Data reduction	Data is reduced according to dimensions

There are many challenges and requirements of data mining like controlling various types of data, time complexity of data mining algorithms, certainty of results obtained through data mining, expressions of data mining results, data security and privacy [11].

Data mining can be classified on the basis of different factors. For example, database, knowledge type, technique used and applications adapted. It involves many disciplines like machine learning, neural networks, logic programming, Bayesian learning. According to N. Jain and V. Srivastava [12], data mining has five major elements that are extract data, store data, provide data to IT professionals, analyze data and show data in proper format. S. H. Liao [13] categorized EDM techniques into three areas that are architecture, knowledge and analysis. These three areas are the main parameters with different perspective, having larger impact on data mining.

III. CATEGORIES OF DMT

Data mining techniques are applied with respect to different aspects of data mining as data obtained from different sources can be different and asynchronous. Data mining is a vast field and found in every field and department. So a specific technique or algorithm is applied for specific type of problem to resolve efficiently. There are nine categories of DMT [13], that are discussed below:

1. Information Systems

Information systems provide a bridge between business world and computer science field. Information system has become the most popular field among all other fields.

2. System Optimization

Original term used for system optimization was 'Linear Programming', in the past. System optimization selects the best element from a set of different available elements.

3. Knowledge-based Systems

Knowledge-based systems are the core of artificial intelligence. There are many tools that are artificially intelligent and produce intelligent decisions through justification [14]. Their base is Artificial intelligence. In these systems; scripts, frames and various rules are used to represent knowledge.

4. Modeling

It is a process in software engineering that uses different data modeling techniques to create a data model. A model in software engineering creates ease of implementing software. It is used to understand a complex structure and flow of a system through different perspective. Modeling techniques are used to analyze the data quantitatively.

5. System Architecture Analysis

System architecture analysis uses a conceptual model that explains the structure, views and behavior of a system. Architecture is a formal description and representation of a system structure. It exposes all components of a system that work together to implement the complete system, their relation and behavior impacting the overall system. A system architecture concerns with the internal interface of system components. A system has hardware architecture, operating system architecture, enterprise architecture and software architecture.

6. Algorithm Architecture

An algorithm is defined as a finite list of instructions to solve a problem. Algorithms are used for data processing and calculation. Algorithm is a main factor effecting the time complexity, cost, and efficiency of a system to resolve a real time problem. Steps involved in the development of an algorithm are shown in Fig. 2.

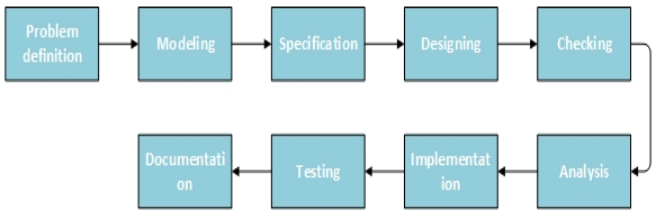

Figure 2: Algorithm development

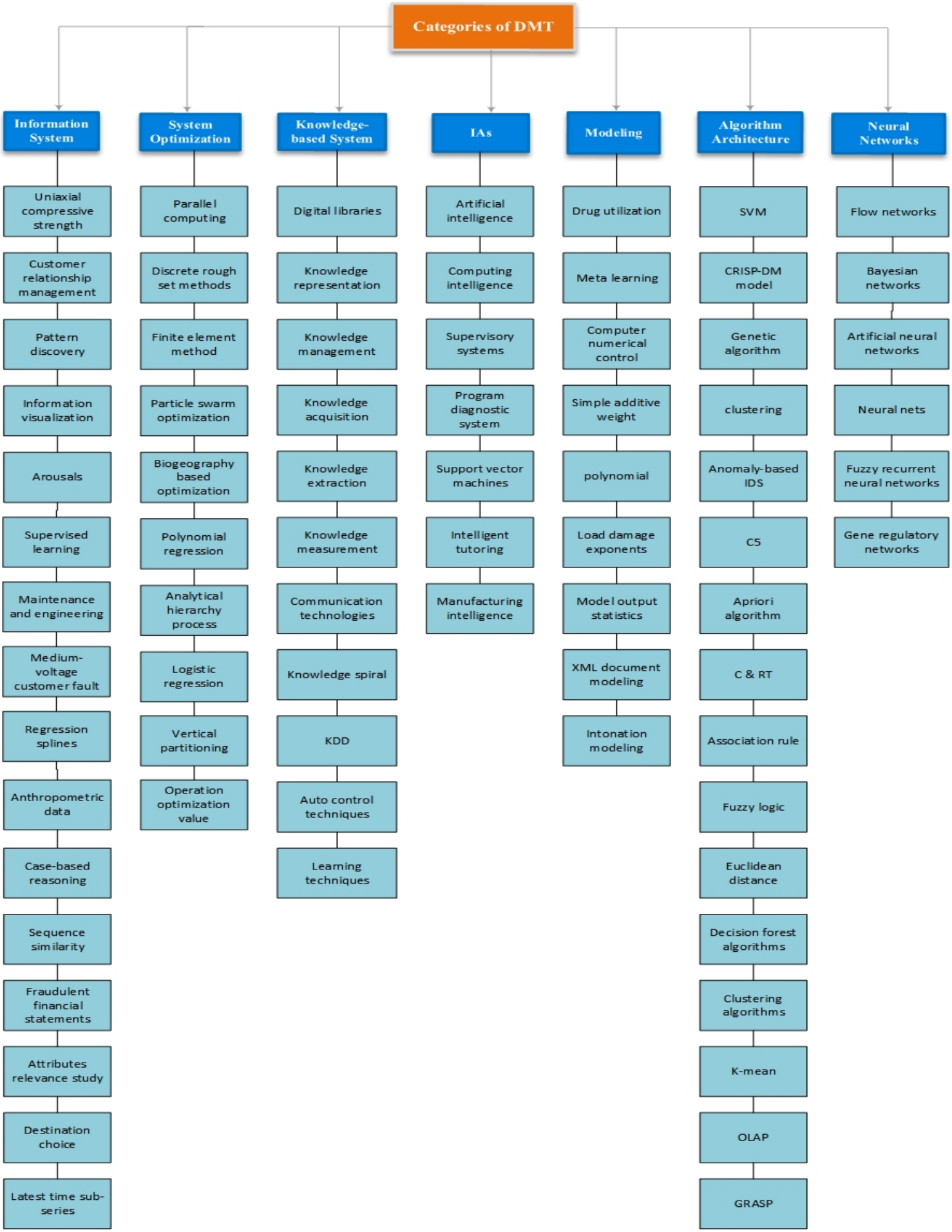

Figure 3: DMT categories

7. Intelligence Agent Systems (IAs)

Intelligent Agent Systems (IAs) are the part of Artificial intelligence, which affect an environment. They can learn and try to achieve their goals using knowledge.

8. Dynamic prediction-based approach

Dynamic prediction-based approach is a mathematical model that is used in modeling molecules. Moreover, it is also used to find the applications in stock market.

9. Neural Networks

Neural networks or artificial neural networks (ANNs) are traditionally referring to biological neuron circuits. Their modern use refers to artificial neural networks that are made of artificial neurons or nodes. They are computational model that are very useful in the field of computer science and many other research fields. Neural networks are composed of neural units or neurons. So each neuron is linked with many other neurons. These systems are not explicitly programmed. However, they are self-learned and trained instead. The goal of ANNs is to work like human brain to solve the problems.

Taxonomy of data mining categories is shown in Fig. 3.

IV. DATA MINING TECHNIQUES

Data mining is a vast field and it has a large number of applications, so it has become interesting subject to research. Data mining techniques are classified into characterization, generalization and association [15]. There are different measures to use data mining techniques as the use of data mining is tricky but helpful if properly used [16]. Some data mining methods are classified and briefly discussed below:

A. Clustering

Data can be in physical or digital form which is stored as big data. Diverse repositories are used to store such data. A data set exceeding the computational limit of software can be termed as big data. In clustering, groups of different objects and their classes are made on the basis of their different aspects like location; connection etc. For example, schools can be grouped on the basis of their similarities or differences. Similarly, students can be clustered on the basis of their behavior. The purpose of clustering is to search data points that are naturally grouped together.

B. Prediction

Prediction often depends on previous knowledge and experience. It is the focus on a single aspect of data with respect to some other aspect of data, called predictor variable. Prediction is used to predict some unknown result on the basis of previous experience or history.

C. Relationship Mining

Relation mining also known as relational data mining is commonly used for relational database. In relationship mining, a relationship is discovered among different variable within a data set. In database, relational data mining algorithm search for pattern among different patterns. Relationship between variables must satisfy two things: interestingness and significance [17].

D. Outlier Detections

Generally, if the new observation is different from the existing one compared, it is named as outlier. Outlier detection compares different values with smallest or largest values in a data set and finds the deviation among values.

E. Text Mining

This data mining technique described as the text data in data mining is specific with text data. Text data include documents, emails, messages, and html files. Text mining can be classified as document processing, document summarization, indexing, topic clustering and mapping [18]. It is commonly used in education and business. Organizations have a big collection of documents and use text mining and obtain information needed [19]. Text mining involves machine learning, statistics and natural language processing. Some of the text mining applications are publishing, telecommunication, information technology (IT), banks, public administration and pharmaceutical companies.

F. Social Network Analysis (SNA)

Social network analysis is a process that uses graph theory and networks to investigate social structures. In this process, relationship between different entities in network information is detected. It is commonly used to analyze the activities of a group or community.

G. Process Mining

Process mining analyzes the business processes on the basis of event log. It extracts the knowledge that is related to process of event log. This event log is noted by the information system for clear representation.

H. Data Distillation for Judgment

In this method, data is represented intelligently. This technique uses visualization and summarization. This is useful to see and explore large amount of data at a time.

V. APPLICATIONS OF DATA MINING METHODS

There are many applications of data mining methods. Some of them are discussed below:

A. Statistics

In the data mining, user of applications is the main subject. Some tools use form usage statistics that are AccessWatch, WebStat, and Analog. One example of usage statistics is measuring the number of visits. If data produce a relational database, then SQL provide many functions such as sample size and mode. All the techniques convert large data into specific visual display. Commonly, large data are described as charts, graphs and 3D representation. These visualized data can be about assignments, exams, courses and marks. Instructors can get information about their students and distance classes.

B. Web data minings

Web data mining is also an application of DM. Here, information is filtered from data obtained from web. Web data include web structure, web content and web usage. The main purpose of web data mining is to facilitate users with information they seek [20].

TABLE 3: Data mining techniques and its applications

Data mining technique / category	Description	Implementation / Tools	Limitation	Applications
Statistics	Graphical and tabular representation of data	WebStat, AccessWatch, Analog	Fail to analyze individual item, Not implemented on heterogeneous data, A small error brings to misleading	Measuring the number of visits on web, Large data is described as charts, graphs, 3D representation
Web data mining	Data mining related to web	Winautomation, import.io, CrawalMonster etc.	Invasion of privacy, Irrelevant contents	Determining the web structure, web contents, web usage
Classification	Grouping of data objects	Generation of groups with same attributes and characteristics	Useless for heterogeneous data	Reduction of information complexity, streamlining in data collection, helpful in planning
Clustering	Grouping of data with samilarities	Cluster 3.0, Java TreeView, PYCLUSTER etc.	Don't support shared storage, operational errors	Fault tolerance, maintenance
Sequential pattern	Ordering of objects with a particular sequence	XAffinity(TM), SPMF, Miningco	Big storage for database,	Shelf in a shop, disaster prediction, proceeding medication detection
Association rule	Antecedent and consequent or if then statements	FPM, Bart Goethals, FrIDA, KNIME, Magnum Opus	Research effort goes to improve the algorithm used, in e-learning, algorithm used has too many parameters	Used in LMS, stock trading
Prediction	Estimation based on previous data	EDM, business	Past result fail due to change in future trend	Weather prediction, student behavior, business
Correlation mining	Creation of patterns from signals, audios, videos, images, sequences	Google Trends, Google Flu Trends, Google Correlate	Doesn't provide the reason of relation among objects	Very helpful for researchers to collect more data than experiments, Used in neuroscience, material science and finance
Casual data mining	Prediction in data relationship	Weka, RapidMiner, KNIME, Rattle	Quality, security, privacy of data	Used in healthcare, business, finance, banking, education
Outlier detection	Detection of deviation of one observation from many other observations	CMSR Data Miner	Need mathematical justification, need probabilistic data model that is complex	Image processing, detection of industrial damage, fraud detection, intrusion detection, inside trading detection, public health and medical
Text mining	Driving information from text	Carrot2, GATE, Gensim, OpenNLP, Orange, Stanbol, KNIME, PLOS, PubGene	A lot of free text in data collection, data is unstructured, syntactic and semantic erros in data, resource development is difficult	Record management, intelligence, social media, searching, publishing, life sciences, security (encryption, decryption), customer relationship management, education, digital humanities
Social network analysis	Use of network to investigate social structures	Commetrix, Cytoscape, Cuttlefish, EgoNet, Gephi, Graph-tool, GraphChi, Graphviz etc.	Risk of fraud, Time wastage, Invasion of privacy	Worldwide connectivity, information sharing, targeted advertising
Decision trees	A tree like model of decisions and their consequences	SilverDicisions, Gambit, Simple Decision Tree, GATree, KNIME, RapidMiner, Smiles, YaDT,	Complexity, loss of innovation, a small change in data set brings a great change in decision trees, difficult to move because of its size and shape etc.	Modeling techniques, feature selection, data preparation, interpretation of data
Nearest neighbor technique	A method used for classification and regression	Face recognition, recommendation engines, spam filtering, Weka, Kaldi, MEKA, mlpy, MODLEM, sgmweka,	Finding the value of k, determining the parameters to be used, high computation cost	Adaptive websites, bioinformatics, cheminformatics, game playing, computer vision, marketing, medical, economics, search engines, stock market analysis, information retrieval, speech recognition
Process mining	Process management on the basis of event logs	Prom, XESame, OpenEXS, ProMimport, MXMLib etc.	Timing problem, conFiguration issues, noise, incompleteness, complexity	Process discovery, conformance checking, compliance checking,

According to Paul B [21], classification technique is used to

- Select students with same characteristics
- Find student misuse
- Find student who are hint-driven in multiple choice questions

In the data mining, common web mining techniques are clustering, classification, text mining, association rule, outlier detection and sequential pattern. These are briefly discussed below:

A. Classification and clustering

Classification and clustering are almost defined the same. Clustering make groups of pages with same contents or users. Classification characterizes the group of user profile and course sessions.

B. Sequential patterns and association rules

Association rules show relation of attributes of a dataset with each other. This relation among attributes creates if-then statements. Sequential patterns tell that which content gives access to the other content.

Michael J. Shaw applied data mining in marketing to support marketing [22]. Sung Ho Ha applied data mining tools in hotel data mart [23]. Usama Fayyad interprets data mining in fraud detection, marketing, manufacturing and telecommunication [24]. Fraud can be associated with different fields like bank fraud, securities fraud, commodities fraud, insurance fraud and any other financial fraud. These all types of fraud are related to financial fraud [25]. Fraud and other related crimes can be categorized to solve and detect through data mining [26]. The firm issuing the fraudulent financial statements can be detected through data mining [27]. Data mining is applied almost in every industry and field. Mrs. Bharati collected some challenges that were faced by some companies in USA, and gave the solutions through data mining [28]. Data mining is beneficial for business as it reduces the cost of business, improves the profit, and enhances the quality [29]. Data mining answers many question of business, which were very difficult and time consuming in the past [30]. It solves many issues of medical field [31]. Nada Lavrac analyzed the medical data using data mining [32]. B. D. Pitt used data mining to load profiling [33]. Andrew M. Wilson elaborated the usage of data mining in pharmacovigilance [34]. Data mining techniques are applied to detect heart disease intelligently [35]. Two or more than two data mining techniques are merged to get better result. Kesheng Wang applied different data mining techniques in enterprise manufacturing process [36]. Data mining techniques are the valuable parameters in chronic kidney diseases [37]. Data mining methods are applied by Sunith Bandaru for knowledge discovery in multi-objective optimization [38]. To discover risk areas on satellite images, data mining techniques are implemented [39].

Data mining techniques, limitations and applications are shown in Table 3.

VI. LIMITATIONS AND OPEN ISSUES

As discussed before, data mining is an important entity, useful in every field. It is categorized and classified into different aspects. We surveyed data mining, its techniques, categories and applications. It is a vast field used by each individual [40]. Learners or students use data mining to personalize e-learning. Better learning experiences are suggested by the data mining. Educators or instructors use data mining to get feedback about instructions. They analyze student's behavior and learning. They also predict the performance of students to improve the customization of courses. Researchers use data mining to choose the best data mining technique to develop data mining tools for specific purpose. Data mining is used by organization and companies to enhance the efficiency in the decision making process. Administration use data mining to find the best way to manage resources and to utilize the resources in the most efficient way.

Data mining has not only advantages, there are also some drawbacks. The main problem in data mining is quality, security and privacy of data [41]. These issues are growing day by day, and researchers are trying to improve and make data efficient by applying proper algorithm. In future, data mining should perceive the complex inputs from the users and should generate the useful and desired results [42].

VII. CONCLUSION

Data being the core entity in every field needs to be managed in efficient way. Data mining helps a lot in this regard. The main issue faced today, is data privacy and data security. In case of global data sharing, privacy becomes more important, especially for web. Therefore, our future work includes the data privacy and security by applying a specific security algorithm that would not harm the data efficiency.

REFERENCES

[1] "Spatial data mining and geographic knowledge discovery—An introduction," *Comput. Environ. Urban Syst.*, vol. 33, no. 6, pp. 403–408, 2009.

[2] M. Goebel and L. Gruenwald, "A Survey of Data Mining and Knowledge Discovery Software Tools," *SIGKDD Explor. Newsl.*, vol. 1, no. 1, pp. 20–33, 1999.

[3] A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," vol. 3536, no. c, 2017.

[4] H. M. Nagy, W. M. Aly, and O. F. Hegazy, "An Educational Data Mining System for Advising Higher Education Students," vol. 7, no. 10, pp. 622–626, 2013.

[5] P. Ristoski and H. Paulheim, "Web Semantics : Science , Services and Agents on the World Wide Web Semantic Web in data mining and knowledge discovery : A comprehensive survey," vol. 36, pp. 1–22, 2016.

[6] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," vol. 18, no. 2, pp. 1153–1176, 2016.

[7] A. K. Choudhary, J. A. Harding, and M. K. Tiwari, "Data mining in manufacturing: A review based on the kind of knowledge," *J. Intell. Manuf.*, vol. 20, no. 5, pp. 501–521, 2009.

[8] C.-L. Huang, M.-C. Chen, and C.-J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Syst. Appl.*, vol. 33, no. 4, pp. 847–856, 2007.

[9] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining," *ACM SIGKDD Explor. Newsl.*, vol. 1, no. 2, pp. 12–23, 2000.

[10] B. Krawczyk and F. Herrera, "Neurocomputing A survey on data preprocessing for data stream mining : Current status and future directions," vol. 239, pp. 39–57, 2017.

[11] M.-S. Ming-Syan Chen, J. Jiawei Han, and P. S. Yu, "Data mining:

an overview from a database perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 6, pp. 866–883, 1996.

[12] N. Jain and V. Srivastava, "Data Mining Techniques: a Survey Paper," *IJRET Int. J. Res. Eng. Technol.*, vol. 2, no. 11, pp. 116–119, 2013.

[13] S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao, "Data mining techniques and applications – A decade review from 2000 to 2011," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303–11311, 2012.

[14] C. F. Chien and L. F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 280–290, 2008.

[15] J. Han and J. Han, "Data mining techniques," in *Proceedings of the 1996 ACM SIGMOD international conference on Management of data - SIGMOD '96*, 1996, vol. 25, no. 2, p. 545.

[16] L. Geng and H. J. Hamilton, "Interestingness measures for data mining," *ACM Comput. Surv.*, vol. 38, no. 3, pp. 1–32, 2006.

[17] R. Baker, "Data mining for education," *Int. Encycl. Educ.*, 2010.

[18] Y.-H. Tseng, C.-J. Lin, and Y.-I. Lin, "Text mining techniques for patent analysis," *Inf. Process. Manag.*, vol. 43, no. 5, pp. 1216–1247, 2007.

[19] W. Fan, L. Wallace, S. Rich, and Z. Zhang, "Tapping the power of text mining," *Commun. ACM*, vol. 49, no. 9, pp. 76–82, Sep. 2006.

[20] R. Iváncsy and I. Vajk, "Frequent pattern mining in web log data," *Acta Polytech. Hungarica*, vol. 3, no. 1, pp. 77–90, 2006.

[21] P. Baepler and C. Murdoch, "Academic Analytics and Data Mining in Higher Education," *Int. J. Scholarsh. Teach. Learn.*, vol. 4, no. 2, Jul. 2010.

[22] M. J. Shaw, C. Subramaniam, G. W. Tan, and M. E. Welge, "Knowledge management and data mining for marketing," *Decis. Support Syst.*, vol. 31, no. 1, pp. 127–137, 2001.

[23] S. H. Ha and S. C. Park, "Application of data mining tools to hotel data mart on the Intranet for database marketing," *Expert Syst. Appl.*, vol. 15, no. 1, pp. 1–31, 1998.

[24] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Mag.*, vol. 17, no. 3, p. 37, 1996.

[25] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011.

[26] S. Wang, "A comprehensive survey of data mining-based accounting-fraud detection research," *2010 Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2010*, vol. 1, pp. 50–53, 2010.

[27] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data Mining techniques for the detection of fraudulent financial statements," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 995–1003, 2007.

[28] M. Ramageri, "Data Mining Techniques and Applications," *Indian J. Comput. Sci. Eng.*, vol. 1, no. 4, pp. 301–305, 2010.

[29] C. Apte, B. Liu, E. P. D. Pednault, and P. Smyth, "Business applications of data mining," *Commun. ACM*, vol. 45, no. 8, pp. 49–53, Aug. 2002.

[30] C. Yen and H. Wang, "Applying data mining to telecom churn," vol. 31, pp. 515–524, 2006.

[31] R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines," *Int. J. Med. Inform.*, vol. 77, no. 2, pp. 81–97, 2008.

[32] "Selected techniques for data mining in medicine," *Artif. Intell. Med.*, vol. 16, no. 1, pp. 3–23, 1999.

[33] B. D. Pitt and D. S. Kitschen, "Application of data mining techniques to load profiling," in *Proceedings of the 21st International Conference on Power Industry Computer Applications. Connecting Utilities. PICA 99. To the Millennium and Beyond (Cat. No.99CH36351)*, 1999, pp. 131–136.

[34] A. M. Wilson, L. Thabane, and A. Holbrook, "Application of data mining techniques in pharmacovigilance," *Br. J. Clin. Pharmacol.*, vol. 57, no. 2, pp. 127–134, Sep. 2003.

[35] S. Palaniappan and R. Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques," *IEEE/ACS Int. Conf. Comput. Syst. Appl.*, pp. 108–115, 2008.

[36] K. Wang, "Applying data mining to manufacturing: the nature and implications," *J. Intell. Manuf.*, vol. 18, no. 4, pp. 487–495, Jul. 2007.

[37] S. Tahmasebian, M. Ghazisaeedi, M. Langarizadeh, and M. Mokhtaran, "Applying data mining techniques to determine important parameters in chronic kidney disease and the relations of these parameters to each other," vol. 6, no. 2, pp. 83–87, 2017.

[38] S. Bandaru, A. H. C. Ng, and K. Deb, "Data mining methods for knowledge discovery in multi-objective optimization : Part A - Survey," *Expert Syst. Appl.*, vol. 70, pp. 139–159, 2017.

[39] B. Boubacar, B. Kamsu-foguem, and F. Tangara, "Data mining techniques on satellite images for discovery of risk areas," *Expert Syst. Appl.*, vol. 72, pp. 443–456, 2017.

[40] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010.

[41] V. S. Sheng, F. Provost, and P. G. Ipeirotis, "Get another label? improving data quality and data mining using multiple, noisy labelers," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, 2008, p. 614.

[42] H.-P. Kriegel, K. M. Borgwardt, P. Kröger, A. Pryakhin, M. Schubert, and A. Zimek, "Future trends in data mining," *Data Min. Knowl. Discov.*, vol. 15, no. 1, pp. 87–97, Jul. 2007.