

# A review of Data Mining Techniques, Applications and Challenges

Rajwinder Kaur

## ARTICLE DETAILS

### Article History

Published Online: 20 January 2019

### Keywords

Data mining, Association rules mining, data, knowledge, Knowledge Discovery in Database etc.

### \*Corresponding Author

Email: rajwinder451@gmail.com

## ABSTRACT

Data mining techniques (DMT) are utilized to change raw data to valuable information or knowledge. Data itself is nothing, yet to process it, is extremely valuable and interesting. Knowledge Discovery in Database (KDD) is the procedure of required yield extraction in various organizations from raw data. Data mining is certainly not a solitary subject. It is a consequence of conjunction of many between disciplinary subjects, for example, statistics, machine learning, neural networks, information retrieval, database technology, and so forth. Association rule mining in huge databases is a significant data mining system and it is utilized broadly in the field of knowledge revelation.

## 1. Introduction

Data mining techniques (DMT) are utilized to change raw data to valuable information or knowledge. Data itself is nothing, yet to process it, is extremely valuable and interesting. There are many development innovations that utilization data as helpful Information intelligently for instance, Knowledge

Discovery in Database (KDD) is the procedure of required yield extraction in various organizations from raw data. KDD is likewise defined as the procedure to see helpful patterns in data. A conventional and most basic outline of data mining or KDD is appeared in Fig.1.

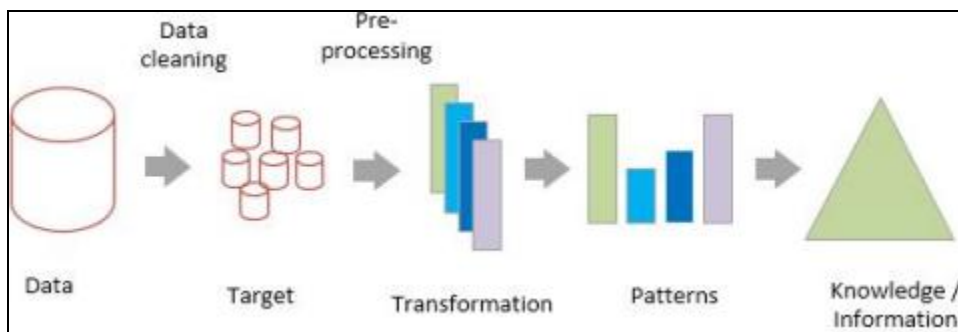


Figure 1: Knowledge discovery from raw data

EDM is the significant utilization of DMT. It is utilized to create such datasets which mechanize the decision of understudies and administrators. It is referenced as an emerging system in training which investigates numerous sorts of data delivered by the educational institutions. It dissects the data produced through the educational framework, and improves the learning and educational impacts. It has a place with the writing that is identified with data mining, visualization, machine learning and calculation. Moreover, methods utilized in machine learning are Naïve Bayes, Neural Networks, K-Nearest Neighbor, Decision Trees and numerous others. Other fields are likewise converged with data mining. For instance, there are many proposed approaches that are the combination of data mining and semantic web. So also, data mining techniques alongside machine learning are utilized in numerous applications.

mining methods are characterized and quickly talked about beneath:

### 1.1 Data mining techniques

Data mining is a huge field and it has countless applications, so it has gotten interesting subject to investigate. Data mining techniques are ordered into portrayal, speculation and association. There are various measures to utilize data mining techniques as the utilization of data mining is dubious however accommodating if appropriately utilized. A few data

- i. **Clustering:** Data can be in physical or computerized structure which is put away as large data. Different vaults are utilized to store such data. A data set exceeding the computational furthest reaches of software can be named as large data. In clustering, gatherings of various items and their classes are made based on their various angles like area; association and so forth. For instance, schools can be gathered based on their likenesses or contrasts. So also, understudies can be grouped based on their conduct. The motivation behind clustering is to look through data points that are normally gathered.
- ii. **Prediction:** Prediction often relies upon past knowledge and experience. It is the emphasis on a single part of data as for some other part of data, called predictor variable. Prediction is utilized to anticipate some obscure outcome based on past understanding or history.
- iii. **Relationship Mining:** Relation mining otherwise called social data mining is usually utilized for social database. In relationship mining, a relationship is discovered among various variable within a data set. In database, social data mining calculation look for

design among changed patterns. Relationship between variables must fulfill two things: interestingness and significance.

- iv. **Outlier Detections:** Generally, if the new perception is not quite the same as the existing one looked at, it is named as outlier. Outlier detection contrasts various qualities and littlest or biggest qualities in a data set and finds the deviation among values.
- v. **Text Mining:** This data mining method depicted as the text data in data mining is explicit with text data. Text data include records, messages, messages, and html documents. Text mining can be delegated record processing, report synopsis, indexing, subject clustering and mapping. It is ordinarily utilized in training and business. Associations have a major assortment of records and use text mining and obtain information required. Text mining involves machine learning, statistics and normal language processing. A portion of the text mining applications are publishing, media transmission, information technology (IT), banks, public administration and pharmaceutical companies.

### 1.1.1 Applications of data mining methods

There are numerous applications of data mining methods. Some of them are examined underneath:

- **Statistics:** In the data mining, user of applications is the main subject. A few tools use structure utilization statistics that are AccessWatch, WebStat, and Analog. One case of use statistics is measuring the quantity of visits. On the off chance that data produce a social database, and then SQL give numerous capacities, for example, test size and mode. Every one of the techniques converts enormous data into explicit visual presentation. Usually, huge data are depicted as outlines, diagrams and 3D portrayal. These pictured data can be about assignments, tests, courses and checks. Instructors can get information about their understudies and separation classes.
- **Web data minings:** Web data mining is additionally a utilization of DM. Here, information is separated from data obtained from web. Web data include web structure, web substance and web use. The main motivation behind web data mining is to encourage users with information they look for.

### 1.1.2 Challenges of Data Mining

There are numerous challenges which can be seen as bottlenecks in the advancement of data mining techniques. Among them, a portion of the main challenges are given beneath.

- **Larye data set and high dimension:** Data mining algorithms need to manage enormous measure of data - both as far as size and dimension. That is the reason, quicker and productive algorithms-are required to handle these tremendous data. Some potential arrangements are sampling, partitioning, parallel processing, and so forth.
- **User interaction and prior knowledge:** Data mining is an interactive and iterative procedure. Here, user's interactions at different stages are required. Domain

knowledge might be utilized either in the type of high level determination of the model or at the more point by point level.

- **Over-fitting and assessing the statistical significance:** Data sets utilized for data mining are gathered from different sources resulting in the misleading data sets. Therefore, a few kinds of regularization methods and sampling techniques are utilized to plan the models for data mining.

### 1.2 Association Rule Mining

Association rules are on the off chance that explanations that help to show the likelihood of relationships between data things within enormous data sets in different sorts of databases. Association rule mining has various applications and is broadly used to help discover sales connections in value-based data or in medicinal data sets. Association rules were first talked about by Rakesh Agrawal, Tomasz Imieliński and Arun Swami in 1993. It is often alluded to as market-basket problem. Association rules find influence of one set of things/properties over another set of items/attributes in a database of exchanges. An association rule has 2 sections:

- an antecedent (if) and
- a consequent (then)

An antecedent is something that is found in data, and a resulting is a thing that is found in combination with the antecedent. Examine this rule for instance:

"If a customer buys bread, he's 70% likely of buying milk."

1. Product (X, bread) → Product(X, spread). Here, things are the things purchased by the customers and an exchange is the things purchased together. The meaning of the rule is that customers for the most part purchase bread and spread together.
2. Age (X, 20-30), income(X, 10000 - 15000) → product(X, mobile). Here, things are estimations of the dimensions from a data distribution center with three dimensions - age, income and product. The rule says that customers with age somewhere in the range of 20 and 30 years and income between Rs. 10,000 and 15,000 will in general purchase a mobile telephone.

Association rules are assessed by the measures, for example, bolster tally, certainty, interestingness, and so on. Along these lines, it tends to be seen as multi-target problem. In any case, it is seen as single-target problem in a large portion of the applications. To the extent applications are concerned, association rules have various applications in different domains. The significant applications are highlighted in the following subsection.

#### 1.2.1 Some Applications

- **Market Basket Analysis:** Nowadays, it is extremely fundamental for the retailers to know the buying inclinations and buying patterns of the customers. On the off chance that a retailer realizes the buying patterns of the customers of a locale, he can detail techniques to pull in the customers by giving proper gifts with various products. Buying patterns likewise help a retailer compose the products in the racks with the goal that customers find the related products together. Along these lines, association rules can assist retailers with getting the buying patterns,

inclinations of the customers, which in turn will increase sales.

- **Financial Services:** Association rule mining assumes a major job in financial segment. Financial specialists use association rules to create investment models, hazard models in stock markets, and so forth. Associating rule mining frameworks have been utilized effectively in stock selection, claims processing by insurance companies, money trading, and so on.
- **Fraud Detection:** With the increase of the utilization of electronic cash like check cards, charge cards, and so on, fraud detection has been one of the prime errands of the crime branches. Association rule mining can find using patterns of cards via card holders, and crime branches can utilize these patterns to distinguish the frauds.

### 1.3 Uses of association rules in data mining

In data mining, association rules are valuable for analyzing and predicting client conduct. They have a significant impact in client analytics, market basket analysis, product clustering, and list structure and store format. Software engineers use association rules to manufacture programs equipped for machine learning. Machine learning is a kind of artificial intelligence (AI) that looks to assemble programs with the capacity to turn out to be progressively effective without being unequivocally customized.

## 2. Literature Review

**Oluwaseun, Adelaja & Chaubey, Mani (2019)** Data mining includes the looking of huge information of the data or records to find patterns and use these patterns in the prediction the future occasions. In most educational segments, for example, secondary schools, polytechnics and colleges; classification system is an indispensable explanatory instrument in prediction of different degrees of exactness. Classification is one of the techniques in data mining for classifying a specific gathering of things to focused gatherings. Fundamental objective of classification is to foresee the idea of a things or data dependent on the accessible classes of things. Development of the classification model constantly characterized by the accessible preparing data set. In this paper we will just examine about the classification algorithms, despite the fact that there are various kinds of algorithms accessible in data mining for the prediction of things to come methodology for a business. The decision tree classification strategy used in this work concentrated fundamentally on data of the understudy's exhibition got in a secondary school during a test utilizing the KNIME tool.

**Tran, Hieu. (2019)** Machine learning and data mining are inquire about zones of computer science whose brisk improvement is because of the advances in data investigation explore, development in the database business and the subsequent market requirements for strategies that are fit for extricating important knowledge from huge data stores. An immense measure of research work has been done in the media region, focusing on various parts of data analytics, for example, the capture, storage, indexing, mining, and retrieval of mixed media huge data. Nonetheless, almost no exploration work gives a total study of the entire pine-line of the strategies utilized in machine learning and data mining in the examination

problems. In this study paper, we lead a complete review of the cutting edge techniques, algorithms machine learning and data mining in sight and sound frameworks.

**Aldowah, Hanan and Al-Samarraie, Hosam and Fauzy, Wan (2019)** the potential impact of data mining analytics on the understudies' learning procedures and results has been acknowledged in advanced education. Consequently, an exhaustive survey of educational data mining (EDM) and learning analytics (LA) in advanced education was directed. This audit secured the most pertinent examinations identified with four fundamental measurements: computer-supported learning analytics (CSLA), computer-supported predictive analytics (CSPA), computer-supported behavioral analytics (CSBA), and computer-supported representation analytics (CSVA) from 2000 till 2017. The significant EDM and LA strategies were distinguished and thought about over these measurements. In light of the aftereffects of 402 examinations, it was discovered that particular EDM and LA methods could offer the best methods for taking care of certain learning problems. Applying EDM and LA in advanced education can be valuable in building up an understudy centered procedure and giving the necessary tools those establishments will have the option to use for the motivations behind nonstop improvement.

**Tummala, Yashasree and Hemantha, Kumar and Kalluri, Hemantha kumar (2018)** the hour of colossal information is directly advancing. Nevertheless, the standard information examinations will in all probability be notable wrench such tremendous measures of information. The request that develops now is the best approach to develop a world class stage to adequately look at immense information and how to design a reasonable mining estimation to find the accommodating things from tremendous information. To significantly discuss this issue, this paper begins with a compact introduction to information examination, trailed by the trades of colossal information assessment.

**Monelli, Ayyavaraiah. (2018)** the field of data mining has seen quick walks in the course of recent decades, particularly from the viewpoint of the computer science network. While data examination has been contemplated broadly in the ordinary field of likelihood and statistics, data mining is a term instituted by the computer science-arranged network. For computer scientists, issues, for example, adaptability, convenience, and computational usage are critical. The rise of data science as an order requires the advancement of a book that goes past the customary focal point of books on just the key data mining courses. Ongoing years have seen the development of the expected set of responsibilities of "data scientists," who attempt to gather knowledge from huge measures of data. In ordinary applications, the data types are so heterogeneous and differing that the principal strategies talked about for a multidimensional data type may not be successful.

## 3. Conclusion

Data mining is certainly not a solitary subject. It is a consequence of conjunction of many between disciplinary subjects, for example, statistics, machine learning, neural networks, information retrieval, database technology, and so forth. Association rule mining in huge databases is a significant data mining system and it is utilized broadly in the field of knowledge revelation. Data mining is definitely not a basic undertaking it is possible that; it is an exceptionally testing

assignment. A portion of the fundamental difficulties of data mining are enormous data set and high measurement, client cooperation and earlier knowledge, over-fitting and surveying the factual criticalness, understanding the patterns, non-standard and fragmented data, blended media data, the board of changing data and knowledge, integration, and so on. Association rule mining is one kind of data mining strategies

and furthermore known as market-basket problem. A portion of the spaces where association rules have been utilized effectively are business, engineering, medicine, telecommunications, and so forth. Some more uses of association rules are: market basket investigation, budgetary administrations, fraud detection, fractional classifications, understanding customers' buying patterns, and so on.

## References

- [1] Oluwaseun, Adelaja & Chaubey, Mani (2019) Data mining classification techniques on the analysis of student's performance 7 17. 10.11216/gsj.2019.04.19671.
- [2] Tran, Hieu. (2019). Survey of machine learning and data mining techniques in multimedia system 10.13140/RG.2.2.20395.49446/1.
- [3] Aldowah, Hanan & Al-Samarraie, Hosam & Fauzy, Wan. (2019). Educational Data Mining and Learning Analytics for 21st century higher education: A Review and Synthesis. *Telematics and Informatics*. 10.1016/j.tele.2019.01.007.
- [4] Tummala, Yashasree & Hemantha, Kumar & Kalluri, Hemantha kumar. (2018). A review on Data Mining & Big Data Analytics. *International Journal of Engineering & Technology*. 7. 92-94. 10.14419/ijet.v7i4.24.21863.
- [5] Monelli, Ayyavaraiah. (2018). Basic Concepts of Data Mining.
- [6] "Spatial data mining and geographic knowledge discovery—an introduction," *Comput. Environ. Urban Syst.*, vol. 33, no. 6, pp. 403–408, 2009.
- [7] A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," vol. 3536, no. c, 2017.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.