International Scholars Journals

African Journal of Library and Information Science ISSN 2756-3383, Vol. 8 (11), pp. 001-006, November, 2022. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Data Mining Roles in Extracting the Knowledge

Eissa Mohammed Ali Qhal[1] & Mohammed Saleh Altowairqi[2]

[1]Jazan University - Applied College - Saudi Arabia.
[2]Information science, King Abdulaziz University, Saudi Arabia.

Accepted 12 December, 2022

Abstract

The data is everything, the data mining is the process of acquiring data from a huge dataset or a database, it can be any source. The knowledge extraction is part of data mining, after acquiring data, the knowledge is acquired through successive completion of other steps like cleaning and integration of data, selection, transformation, pattern evaluation and knowledge extraction. The motive of the research is to summarize the existing researches and studies on the role of data mining in the process of knowledge extraction. To identify the research papers, the author conducted a rigorous data collection from all the resources. The researcher accessed many research papers from various publishers, and found some relevant studies such as data mining and analytics in the process industry, the role of machine learning, combining web data extraction and data mining techniques to discover knowledge, role of data mining techniques, a review on knowledge extraction for business operations using data mining, applying data mining techniques for descriptive phrase extraction in digital document collections and data mining techniques and application. The study reveals the motive, algorithms and their uses, findings, scope of the studies of the recent studies and how these studies are used.

Keywords: Data mining, Knowledge extraction, algorithms, analysis, extraction, data mining techniques.

I. INTRODUCTION

The word “data mining” is called the method which offers with the distillation or elimination of unseen predictive information from massive database. It consists of one of a kind categorization of statistics through massive quantities of statistics units and find out beneficial and vital data from it. It is commonly carried out within the discipline of commercial enterprise, and additionally performs a primary position within the discipline of finance, however it is most crucial within the discipline of science on this region it's far used to filter precious facts from massive statistics units produced via way of means of modern day experimentations, observational methods and strategies [12]. The process of data mining is shown in figure 1.

It is visible as a step by step crucial device to convert statistics into business intelligence in order to achieve beneficial result. Data mining is generally the project of locating advanced innovative and unusual sequence or patterns from saved statistics[1]. It generally determines the sheer within the information that's past the easy evaluation with the assist of algorithm. Data mining may be obligatory on approaches of knowledge discovery in addition to prediction [18]. Various applications of data mining is shown in figure 2.

The feature exploration of conduct, correlations, patterns and trends, permitting the proper choices to be evaluated and brought on the proper time, and suitable answers to planning, problems, modernization and improvement in all fields[3]. Examination procedures answer many questions, and in record time, especially those types of questions that have been tough to respond, if now no longer incredible, the usage of classical statistical evaluation performances, which, if any, take a long term and lots of evaluation procedures [5][6][7]. There are diverse groupings in the process of knowledge extraction as follows:

- **Characterization**: Data characterization is a summarization of the overall traits or capabilities of an objective elegance of information. The information similar to the user-distinct elegance are normally accrued via way of means of a question. For example, to observe the traits of software program merchandise with income that accelerated

Corresponding author E-mail: egohal@jazanu.edu.sa

Figure1: Data mining process.

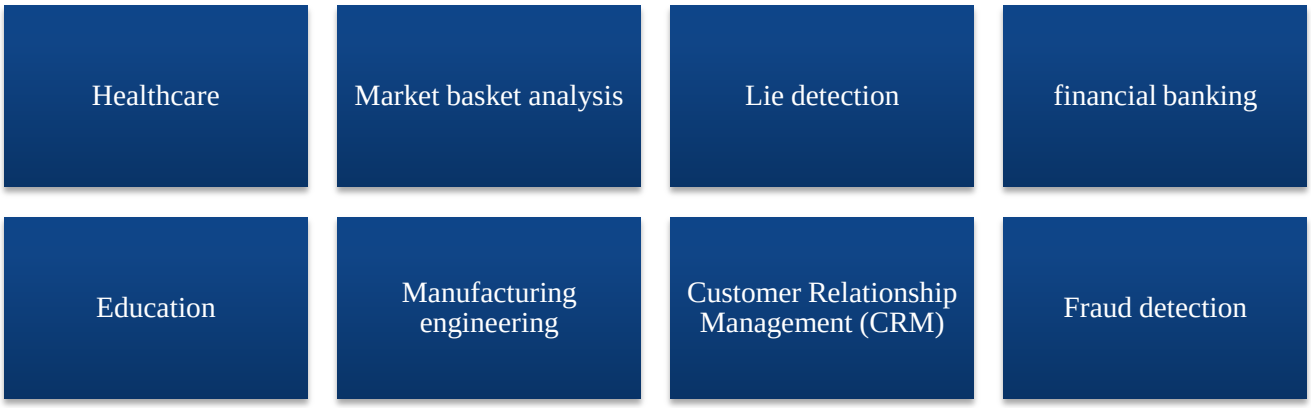

Figure2: Applications of data mining.

via way of means of ten times within the preceding year, the information associated with such merchandise may be accrued via way of means of executing an SQL query at the income database [14].

- **Discrimination**: The term data discrimination defines the separation or the categorisation of data using the appropriate effective algorithms and techniques. The term called “Data Discrimination” can be also called “discrimination by algorithm” this is a proven concept which happens when the predefined data source or the data types would purposely or involuntarily be executed specifically compared to others [1].
- **Association and Correlation Analysis**: Correlation coefficients offer a numerical size of the affiliation among variables. They may be used to decide the further among gadgets while they're merged right into a cluster; to evaluate the association among gene expression profiles, to set up a connection among genes in a genetic network, or to assess the settlement among experimental procedures [7].
- **Classification**: It perceives the class or the magnificence label of a newly obtained observation. First, a fixed gathering of information is used as education information. The set of entered information and the conforming outcomes are quantified to the set of rules for further operations. So, the training set of data consists of the entered information and their related magnificence tags. By means of the training set of data, the chosen algorithm derives a classifier or the version.
- **Prediction**: It is usually used to discover a numerical outcome. Similar as in classification, the training set of data consists of the user inputs and corresponding numerical outcomes values. The algorithm derives the version or a predictor in line with the training set of data. The version needs to discover a numerical outcome while the brand new information is given [13].
- **Outlier Analysis**: The outlier analysis is a step where the anomalous observations are identified from the dataset. Outlier analysis has numerous applications for medical diagnosis, financial industry, Web analytics, quality control, intrusion detection, and fault diagnosis.
- **Evolution Analysis**: Evolutionary analysis can display the genetic collection dating of class within the evolution system from the attitude of molecular development. In the system of organic data evaluation, phylogenetic trees are regularly used to give the evaluation results [5].

1.1 Data Mining in Extracting the Knowledge

Data Mining is generally utilized by corporations with extreme advertising company, customer demands- Retail, customer preferences, financial, decide price, product positioning, Communication and effect on sales, and client satisfaction, company profits. Data mining permits a store to apply point-of-sale information of client procurements to broaden promotions and merchandise that assist the employer to draw the client.

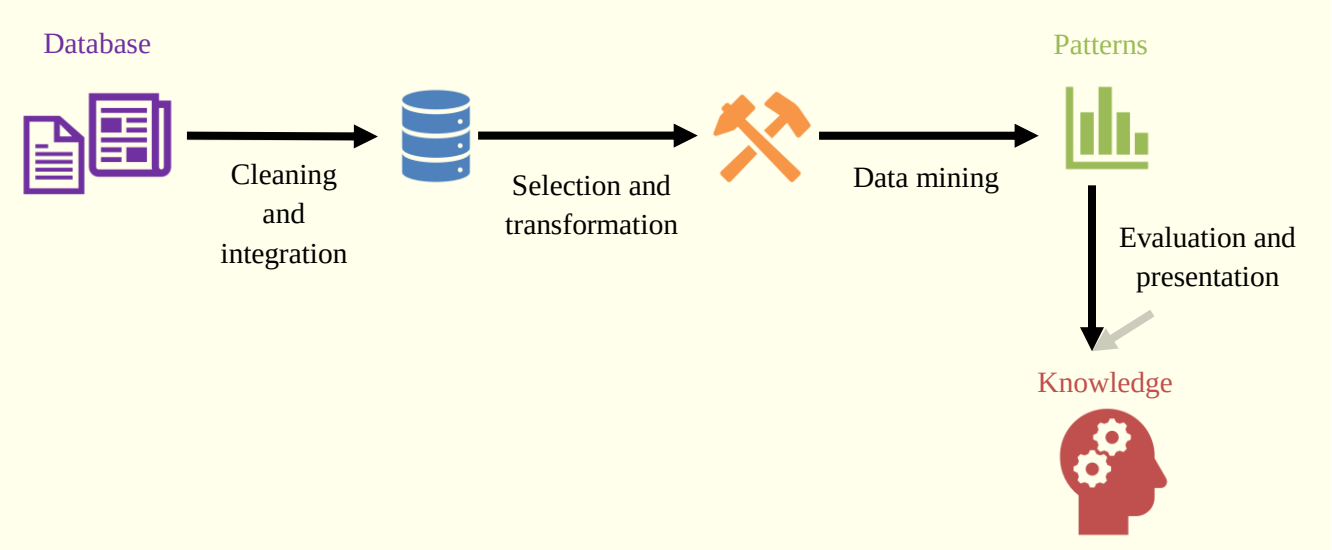

Figure 3: The process of knowledge discovery.

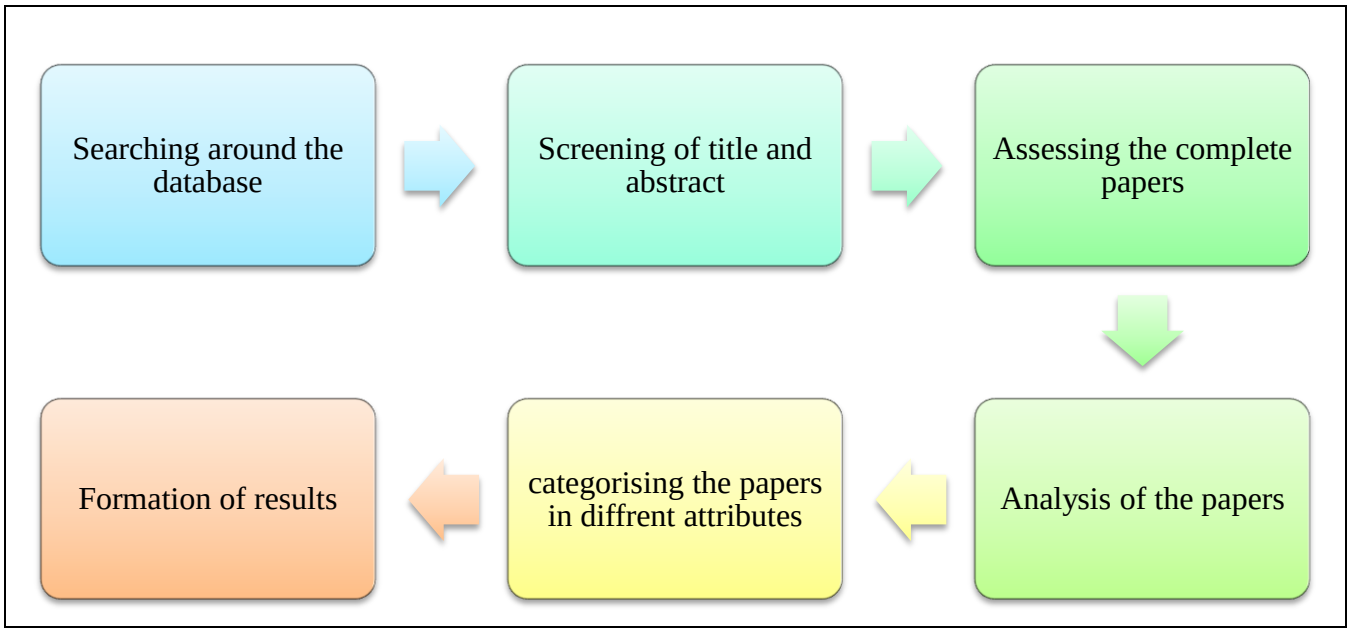

Figure 4: The flow diagram of the study.

Feature Extraction makes use of an object-primarily based totally method to categorize descriptions, wherein an object (additionally known as segment) is a set of picture element with comparable texture attributes and/or spatial, spectral [6]. Traditional type techniques are picture element primarily based totally, which means that spectral statistics in every picture element is used to categorize imagery [17]. The process of knowledge discovery is shown in figure 3.

For everything literally, the data or the dataset or databases are the primary block. The data can be obtained through many ways like survey, research papers, experience and so on. In this study, most of the source are the research paper. From the papers, the author has filtered the data and obtained the results.

II. Literature review

D. Sai Pranav et al. has stated in his research that the cloud computing is now emerging with numerous innovative ideas in association with the concept of data mining. He has also proved that “Data mining in cloud computing is the method of extricating organized data from unstructured or semi-structured web information sources” [15]. T. Rajesh Kumar et al. has proposed a study which reveals an implementation of

Table 1: A review of articles.

S.no	Article	Objective	Algorithm/ Analysis/ Techniques used	Findings/Results/ Conclusions	Future
1	Data mining and analytics in the process industry the role of machine learning	To provide a review on the already existing data analytics and data mining applications from the post decades.	Principle component regression, Partial least squares, fisher discriminate analysis, multivariate linear regression, Artificial neural network, Support vector machine, Nearest neighbor, Gaussian process Regression, Decision tree, Random forest	The ten supervised learning algorithm and eight unsupervised learning algorithms are used to obtain the efficiency of data analytics and data mining algorithms.	Authors are expecting that data analytics and data mining would become an important part in the process industry along with the machine learning technologies.
2	Combining web data extraction and data mining techniques to discover knowledge	To propose a methodology to apply the clustering notion on categorical web data and to use the clustering results as part of the input for the classification conducted on another set of data.	K-means, Naïve bayes, K-mode	The proposed methodology performed effectively on the dataset with the help of their parameters such as precision for the clustering algorithm, accuracy, the error of classification and results.	The authors suggested the use of classification techniques for extracting additional text such as required skills, contract type, business sector and education level considering that this kind of information may be vary relevant for the job searches and recruiting agencies.
3	Role of data mining techniques	To review different data mining techniques which can help researchers to complete their study	Neural network algorithm, Support vector machine, Naïve bayes classification, decision tree	The data mining algorithm to obtain greener and smarter environment are researched and is explained in detail. By using optimized parameters, the large dataset is trained to predict energy consumption and to enhance accuracy.	Author claims that deep learning technologies can be applied to train large quantity of data and dataset.

judging the students' performance through analyzing different types of parameters namely, condition of the class, funds, gender, marks, and more. They have given a clear summary of their study based on clustering and classification algorithms; Web based system, neural networks, nearest neighbor methods, Bayesian classification-mean algorithms

Table 1: cont.

4	A review on knowledge extraction for business operations using data mining	To discover different models of knowledge management and business operation and to study the importance of the data mining techniques.	Data mining techniques: association mining, classification techniques, clustering techniques, text mining	The author found that implementing knowledge management and data mining techniques for business operations would be helpful. During the study they proposed various models and concluded with its implementation.	In future, from the business sector there exist a possibility to get instructions to the human and support to identify the work and determine the solutions.
5	Applying data mining techniques for descriptive phrase extraction in digital document collections	To prove that the general data mining techniques can be implemented in text analysis procedures like descriptive phrase extraction.	Episode rule algorithm, association rule	Presented a general framework for text mining along with general data mining methods. Illustrated some examples of after and before process procedures.	Their proposed methodology can be implemented in the real time problems which can be obtained by proving that episode rules and episode produces differences between files.
6	Data mining techniques and application	To study about the data mining and Knowledge discovery in databases process such as finding evaluation, finding interpretation, finding presentation, pattern searching, data transformation, data cleaning, and data selection.	Knowledge Discovery in Database, association, clustering, trend analysis, machine leaning algorithms	The researcher found that the data mining techniques can be used in the field like statistical approaches, machine learning approaches, database oriented approaches and neural networks.	The data mining applications can be used in the future for various sectors such sports, money laundering monitoring, healthcare management and tax fraud detection.

and Naive bayes approaches to examine replacement implementation. Researchers in lots of unique fields have proven awesome interest in knowledge discovery [8]. Several rising programs in information-supplying provisions, together with data warehousing and on line provisions over the Internet, additionally demand for diverse data mining strategies to higher apprehend person behavior, to enhance the provider furnished and to growth commercial enterprise opportunities[3].

III. METHODOLOGY

Methodology refers back to the overarching method and motive of a research. It includes reading the strategies used on the research and the theories or concepts at the back of them, so as to expand on method that fulfills the objectives of the researcher's current study[2]. The entire flow of the study is shown in figure 4.

The most initial step in every research or any project is to find the sources and the data. The data is an important part in the researches, the data for this study is taken from different relevant studies and publications from different websites and reference books [9]. Then the author took each and every source which was relevant to the study and selected the abstract, through those abstract people can get an idea of the complete research. After selecting abstract, the author accessed the complete paper to analyze what are the major objectives, findings, whether the paper is beneficial in future or did they mention any scope of their researches and what all kind of algorithm and techniques are used and so on[16].

IV. RESULTS AND DISCUSSIONS

As a result, this study has come up with proof that data mining plays an important role in knowledge extraction. As a part of the research the author found some relevant resources which shows the necessity of the data mining techniques [11]. The author found that the researchers have used many algorithms

to figure out their research problem, the most common algorithms are K-means, Neural network, Naïve bayes, decision tree, Support vector machine, classification, clustering, regression, random forest, artificial neural network and more[4]. Comparison of various research articles are shown in table 1.

V. CONCLUSION

As a conclusion, making use of data mining tools, to look for, mine, and discover traits and patterns to discover patterns of conduct and traits, or with the aid of using the usage of prediction and classification that predicts what would possibly appear within the destiny. For instance, the clients' possibilities are discovered after they purchase a few merchandises over different merchandise, and the possibility that clients will purchase precise merchandise is expected consistent with the supply of information that is premeditated on historic information within the database, or, for example, the conduct of clients is discovered after they purchase certain merchandise with different merchandise. For such situation the data mining can be use [10]. With time the data mining will grow rapidly. Within the coming decade everything will be working in data mining techniques. In coming days, the business sector will be driven by data mining tools to get easy decision making and instructions. Suppose in a startup enterprise, they might not be familiar with the current trend, user demands, user experience, user expectation, the real world problems etc. But the data mining can help them, as a result of data mining they will achieve all data needed for their initial steps to start their production later they can perform knowledge extraction to predict and work accordingly.

REFERENCES

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules", Proc. 1994 Int. Conf. Very Large Data Bases (VLDB'94), pp. 487-499, September 1994.

[2] M. Pletikosa, "Primjena metoda umjetne inteligencije na poveÄ‡anje sigurnosti uloga za pristup bazama podataka", July 2010.

[3] R. Sandhu, D.F. Ferraiolo and D.R. Kuhn, "The NIST Model for Role-Based Access Control: Toward a Unified Standard", 5th ACM Workshop Role-Based Access Control, pp. 47-63, July 2000.

[4] Kononenko and M. Kukar, "Machine Learning and Data Mining: Introduction to Principles and Algorithms" in , UK:Horwood Publishing Chichester, 2007.

[5] M. Kuhlmann, D. Shohat and G. Schimpf, "Role mining - revealing business roles for security administration using data mining technology", Proc. of the eighth ACM symposium on Access control models and technologies (SACMAT '03), pp. 179-186.

[6] J. Pei, "Pattern-growth methods for frequent pattern mining", 2002.

[7] J. Vaidya, V. Atluri and J. Warner, "RoleMiner: mining roles using subset enumeration", CCS '06 Proc. of the 13th ACM conference on Computer and communications security. ACM, pp. 144-153.

[8] R. Agrawal, T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases", Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'93), pp. 207-216, May 1993.

[9] J Han, J Pei and M. Kamber, Data mining: concepts and techniques, Elsevier, Jun 2011.

[10] S. Ijaz, M. A. Shah, A. Khan and M. Ahmed, "Smart cities: A survey on security concerns", Int. J. Adv. Comput. Sci. App, vol. 1, no. 7, pp. 612-625, 2016.

[11] F. Wang, N. He1ian, Y. Guo and H. Jin, "A distributed and mobile data mining system In Parallel and Distributed Computing Applications and Technologies", 2003. PDCAT'2003. Proceedings of the Fourth International Conference on, pp. 916-918, 2003, August.

[12] H. Kargupta, B. H. Park, S. Pittie, L. Liu, D. Kushraj and K. Sarkar, "MobiMine: Monitoring the stock market from a PDA", ACM SIGKDD Explorations Newsletter, vol. 3, no. 2, pp. 37-46, 2002.

[13] H. Kargupta, R. Bhargava, K. Liu, M. Powers, P. Blair, S. Bushra, et al., "VEDAS: A mobile and distributed data stream mining system for realtime vehicle monitoring", Proceedings of the 2004 SIAM International Conference on Data Mining, pp. 300-311, 2004, April.

[14] C. Comito, D. Talia and P. Trunfio, "An energy-aware clustering scheme for mobile applications. In Computer and Information Technology (CIT)", 2011 IEEE 11th International Conference on, pp. 15-22, 2011, August.

[15] W. Sherchan, P. P. Jayaraman, S. Krishnaswamy, A. Zaslavsky, S. Loke and A. Sinha, "Using on-the-move mining for mobile crowdsensing", Mobile Data Management (MDM) 2012 IEEE 13th International Conference on, pp. 115-124, 2012, July.

[16] H. Tayebi, S. Krishnaswamy, A. B. Waluyo, A. Sinha and M. M. Gaber, "Ra-sax: resource-aware symbolic aggregate approximation for mobile ECG analysis", Mobile Data Management (MDM) 2011 12th IEEE International Conference on, vol. 1, pp. 289-290, 2011, June.

[17] C. Comito and D. Talia, "Energy consumption of data mining algorithms on mobile phones: Evaluation and prediction", Pervasive and Mobile Computing, vol. 42, pp. 248-264, 2017.

[18] E. J. G. Emberda, L. M. N. Dalagan and C. F. O. Baguio, "Recognition of Traffic Weight Using Sobel Edge Detection Method and K-Nearest Neighbor Algorithm", UIC Research Journal, vol. 18, no. 1, 2012.