TECHNOLOGIES USED IN DATA MINING

Yazeed Al Moaiad[1], Daud Elega[2]

Al-Madinah International University, Malaysia

[1]yazeed.alsayed@mediu.edu.my

[2]daud.elega@mediu.edu.my

Abstract: **By defining data mining, which is an examination of a large database that has been previously archived for the purpose of extracting new data of high value. Powerful technology has great impact and great potential to help companies pay attention to the most meaningful and important data in the repositories in which the data is stored. Hence the role of future data mining tools and techniques, which enable businesses to be more predictive and proactive and provide an advantage in making informed decisions. Hence, data mining techniques can answer questions related to the work of institutions, which may take a long time to find a solution to them. So, this paper aims to evaluate data extraction tools and techniques, e.g., resolution, tree prediction, neural networks, and clustering.**

Keywords: **Data mining, Decision, Classification, Neural Networks, Association.**

I. INTRODUCTION

All companies that care about a brand consider big data to be one of the most important technologies that help the brand grow. In our current era, large companies use very large data analysis tools and techniques to benefit from it in achieving important goals in their interest, for example, but not limited to customer satisfaction, as well as organizational growth. Hence the importance of understanding and analyzing big data for the growth and prosperity of companies and the ambition of expansion. Therefore, mining tools and techniques are of great use to help companies understand and analyze that data in highly effective and effective ways. Hence, these tools and techniques help in solving many problems and provide a contemplative and insightful view of any company. [1]

It is no secret for those interested in techniques and tools for data mining that each sample has a different effect on the outcome. As it was said previously about the brand and its importance to companies, knowing the business problems and the reasons that led to that, will give a shorter time for those companies to properly search for correct and specific data to reach better results. So, these techniques such as resolution, tree prediction, neural networks, and clustering use coded words to describe an entire range for the purpose of analyzing the data. Therefore, it is necessary to consider mining techniques to have an effective impact on the data and to explore records on a typical or regular basis and focus on obtaining accurate and accurate data. Hence the importance of developing specific strategies that help in using data extraction technology in a simple way that the company understands. A tracker of data mining tools and techniques knows that the main goal is to collect specific information, which is applied to larger data sets. [2]. Some mining techniques and data mining tools will be mentioned, which will contribute to the development of the business and push it forward in an efficient and comprehensive manner.

II. DEFINITION OF DATA MINING CONCEPT

A. Definition of Data Mining

Data mining can be defined as a procedure to define a pattern or patterns of data for the purpose of obtaining and finding useful information from a group of enormous data, which ultimately enables a decision or judgment about a specific thing. Several data mining tools and techniques have been developed and have already been used in projects for the purposes of mining specific data related to neural networks, decision tree, prediction, clustering, and classification. These types of exploration technology have specific rules and methods that help determine what problem needs to be solved. In what follows, the data extraction technology will be studied in a specific and brief way.

B. Association

Implementing encryption will protect data from three sets of parties: The definition of association, which is one of the important data mining tools and techniques, lies in exploring patterns based on the relevant relationship between variables in a single transaction. And this relationship that exists between the elements helps to explore the frequency between those elements, and these elements are often different, and therefore it is noticed that those frequencies that are inside a set of certain data will use the if-then correlation rules for the purpose of obtaining the probability of that relationship between all the different elements. , Or any variables that might be within the scope of this set of big data that is in many different types of databases. Association rule mining includes several applications that may be widely used in several areas for the purpose of exploring, for example, sales correlation in transaction data or using other areas such as medical data sets [3].

Hence it is observed that many retailers are using the bond on a more comprehensive scale which helps in understanding customer behavior in buying. Based on the collection of customers' purchasing data, retailers can know who the regular customers are in buying a specific product such as a lighter while buying a pack of cigarettes, so a pack of cigarettes and a lighter can be placed side by side to save some time and increase productivity in sales [4]. Hence the correlation between retail and association mining as a graphical analysis of the market basket.

C. Classification

The definition of classification, which is among the tools and techniques of mining critical data, is to classify and divide a group of big data into a different, smaller set that is more easily predictable. The classification has a clear effect on forming an idea about the customer category, for example, or a specific element in a group of data, by describing and defining multiple characteristics for the purpose of defining a specific category accurately. An example of this: Classifying the building for different buildings according to certain and different characteristics such as height, building structure and unit. Hence, the type of building is known, and the target group is determined according to age, gender, and social class by comparing specific characteristics within the database. [3].

D. Clustering

One of the most important and oldest techniques used in mining and data extraction is clustering. Therefore, clustering is the analysis of a characteristic or several characteristics to determine which data may be similar to each other, for the purpose of understanding the similarities and differences between those previously defined data. The process of clustering is sometimes called segmentation because it segments the data into categories to identify a cluster of correlating results. For example, we can use clustering technique for book management in a library to keep books that have some kinds of similarities in one shelf and give it a meaningful name for readers to grab books in that topic easily, instead of looking for the whole library.

E. Decision Tree

Decision tree techniques can be applied which is one of the criteria by which to select specific data within the overall hierarchy.
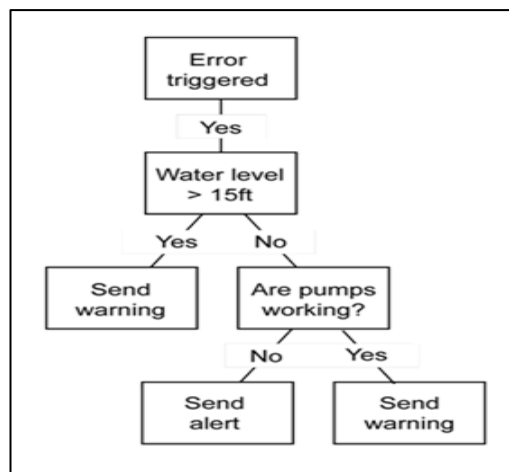

FIGURE 1: DECISION TREE STRUCTURE

The decision tree may advance a question that may result in at least two or more answers supporting a classification or identifying specific data for the purpose of classifying it. Moreover, a prediction can be used based on each answer. Therefore, classification systems that are associated with a decision tree are often used for the purpose of associating type information with predictive systems. Predictive systems may have different predictions that depend mainly on experience, which contributes to distinct outputs and a sophisticated decision tree structure [5].

F. Prediction concept

When talking comprehensively about the subject of anticipation and moving from predicting the components that may be failing to knowing the types of fraud and ending with forecasting and knowing the profits of companies better. The prediction is linked to the techniques of exploration and is used with it in the exploration of other data, and as we know, forecasting helps in analyzing events and trends and the extent of matching patterns, classifications, and relationships. An example of this is analyzing the credit card decision tree, linking it to previous transactions, and matching the classification pattern at a specific time for the purpose of finding out whether the transaction is a form of fraud. For example, matching tickets for flights purchased to a specific country and transactions in that country, through which to know whether the transaction is correct or not. [5].

G. Neural Networks

The concept of the neural network, which is an important technology in terms of use in our time. Often this technology is used in the advanced stages of mining technology to extract data. This neural network was formed from the artificial intelligence community, and due to its largely automation, its use becomes very easy. This gives the user more space, and it is not necessary to know what is related to work or databases [6]. On the other hand, we must know some points regarding the concept of a neural network to work more efficiently.

• How to deliver the contract, how is it done?

• Determine and know the number of processing units required for use

To what extent we can continue training, and when should we stop it.

Neural networks have two main parts:

1. Node: It is a comfortable, freely compatible with the neurons of the human brain

2. Link: It is a comfortable, freely matching of the connexions in between the neurons of the human brain, which form a group of neurons a network or several interconnected networks, which contribute to the formation of a layer or several different and multiple layers. This is a kind of configuration of neurons and their interconnection with each other by network geometry.

And speaking of neural networks, which is an important model for prediction technology, but it is difficult to understand even for experts working in this field. This is because it produces very complex templates that are difficult for experts to understand and analyze clearly, and these networks have many applications, for example, fraud detection processes resulting in a specific transaction in a business [6].

III. CONCLUSION

This paper shows a detailed representation of data mining techniques. Big data is a term of concerning large volumes of complex data sets. Data mining comprises extracting useful rules or interesting patterns from historical data. The high-performance computing paradigm is required to solve the problem of big data using some techniques that list in this paper for example Clustering, Decision, tree Prediction, and Neural Networks. The mentioned data mining techniques have been a guide to lead the business in a comprehensive and successful manner.

REFERENCES

[1] M. Rouse, "association rules (in data mining)," Techtarget, [Online]. Available: https://searchbusinessanalytics. techtarget.com/definition/association-rules-in-data-mining. [Accessed 3 December 2018].

[2] Zentut, "Data Mining Techniques," Zentut.com, [Online]. Available: http://www.zentut.com/data-mining/data-mining-techniques/. [Accessed 3 December 2018].

[3] EDUCBA, "ntroduction to data mining," Educba.com, [Online]. Available: https://www.educba.com/course/introduction-and-applications-of-data-mining. [Accessed 3 December 2018].

[4] EDUCBA, "Data Mining Techniques for Successful Business (Tools, Software)," Educba.com, [Online]. Available: https://www.educba.com/data-mining-techniques/. [Accessed 3 December 2018].

[5] M. Brown, "Data mining techniques," [Online]. Available: https://www.ibm.com/developerworks/library/ba-data-mining-techniques/index.html. [Accessed 3 December 2018].

[6] EDUCBA, "7 Important Data Mining Techniques for Best results," [Online]. Available: https://www.educba.com/7-data-mining-techniques-for-best-results/. [Accessed 5 December 2018].