

# Developing Conception about Differences between OLAP And Data Mining

Dr. Ajay B Gadicha<sup>1\*</sup>, Dr. V. Gokula Krishnan<sup>2</sup>

<sup>1</sup> P R Pote Patil College of Engineering and Management Amravati, India

<sup>2</sup> Professor, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Thandalam, Chennai, Tamil Nadu, India

\*Corresponding Author Email: [ajjugadicha@gmail.com](mailto:ajjugadicha@gmail.com)

---

## Abstract

The following study has been based on the concept of data mining and OLAP and the differences between these aspects. At the beginning, the procedures of data mining and OLAP have been served with a brief introduction. The discussion has focused on the section namely the introductory section, the methodology section, results section, discussion and the conclusion part. The results part has given a detailed account about in ways the data mining is different from the OLAP. Both the technology has been placed separately and finally into result section key difference has been analyzed thoroughly. Data is an emerging technological tool that was going to be a revolutionary in the future technology sector but into the current the use of data mining is complex and it is difficult to make the most as the technology is costly as well as sales expertise is available in the market. Due to this it can be said that the technology is at a developing stage and it needs further improvement. OLAP is the type of technological tool that is often used to make the process more and more convenient with the analysis of past data. Both the technology has its own significance and differences that have been given in the research study.

## Keywords

Analysis, Data Mining, OLAP, Technological.

---

## INTRODUCTION

In the present era of data digitization, data mining is a core source to justify the data in a certain manner and on the other hand, OLAP is another aspect which helps to bring validation in the process of executing the data in a certain manner. The full term of OLAP is online analytical processing which is an example of analysis services. This is technology that intends to arrange large business databases and supports difficult analytical parts for a complex sort of analysis [1]. Also OLAP can be implemented to evaluate difficult and hectic queries related analysis without negatively making an impact on the system of transmission. Generally, OLAP is a type of software technology that permits the users to examine and interpret the insights from several database systems at the same time. It is dependent on the multidimensional data model and permits users to ask questions on multidimensional data in a certain way.

Later on data mining is an explanation of solving the problems about a business in an organized manner. By examining the bonds among the parameters like age of consumers, gender, intention, basically it helps to justify the demographical data of an individual. In short, data mining is the procedure of discovering the knowledge from a raw database. Data mining, also known as knowledge discovery in data, is the procedure of unfolding the paradigm and other source-full insights from a large set of data. There are types of data mining procedure which has been followed on the trait of the database and the types are as follows- Predictive data mining and descriptive data mining [2]. Data mining has been implemented to explore the large database immensely

and to develop market segmentation in a certain manner.

Data mining is the reference to the field of computation which intends to deal with the data which has been extracted from the authentic sources, trends and paradigms from a huge data set. On the other hand, OLAP is a type of technology of immediate entry to data with the help of multidimensional patterns of the dataset [3]. OLAP intends to deal with the summary of data and transactional data in a certain way.

## MATERIALS AND METHODOLOGY

The material and methodology part is the part that is going to discuss the research methodology that has been followed while preparing this research study. The given research topic is based on finding the differences between the OLAP and the Data mining. Both of these have their own significance in their places and this can be understood with the help of this research study. In this research study it has been discussed thoroughly about the data analysis and the way the data has been collected. For this research the data has been collected on the basis of secondary data analysis and the data analysis methods that have been used here is thematic data analysis [4]. In thematic data analysis the data are analyzed on the basis of preparation of themes which are on the basis of the research topic. If the data is analyzed with the help of thematic analysis the data that is ideal for this is secondary data which includes the data from journals, books, government data and the data collected from other theories. Moreover, it can be said that there is a need for the technology of data mining in various fields like in data privacy, data security, studying the current trends and many more.

**RESULTS**

The online analytical processing (OLAP) is an approach for answering multidimensional enquiries in computing. It can be said that the term OLAP is a part of the broad category of business intelligence and for that it encompasses other technology like data mining. Data mining on the other hand can be explained as the machinery that basically turns the raw data into useful information with the careful analysis of the raw data that are provided [5]. There are some of the differences between the OLAP and data mining and those differences are as follows: in data mining multidimensional analysis is done whereas in the OLAP online analytical processing is done. The number of dimensions for the data mining is higher than that of the OLAP. Data mining often deals with the summary of the study whereas the OLAP deals with the detailed transnational data. Today both of these technologies are used for the data analysis and data scores are the subject those explorers about the study.

**Data mining and its uses**

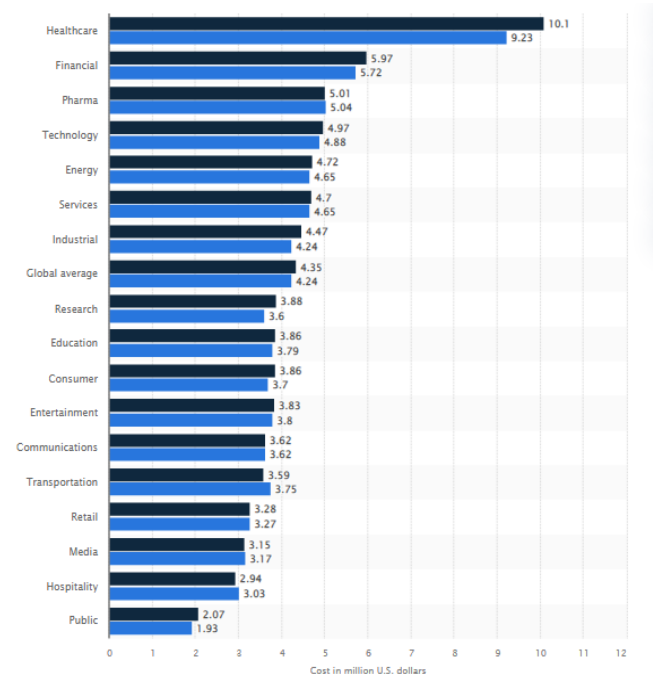
Data mining is the process which involves the extraction of the usable data from the larger set of data which are in the form of raw data. Today data mining is used by a number of companies like Amazon, Netflix, and others to know better about the likes and dislikes of the customers. The technology these days is also used by the companies which work for the protection of online data from the threat of theft. Information security has become the biggest concern for the people of the world and for that there has been a continuous movement in the technological sphere that is known as data mining [6]. There are a range of benefits of data mining and some of the benefits are as follows: it predicts the values that are based on the outcomes that are most likely; also the process of data mining involves analysis of large amounts of data. It has also been said by some of the experts that the process of data mining is automatic as the values are generated automatically.

It can be said that the threat of the cyber-attack has increased in recent years and cybercrime attacks in the various sectors have been given in the figure below.

In this figure globally the data breach has been in the industry and for that there is data mining technology that can be used. The most affected sector due to the data breach is health care. In the year 2020 the total cost of data breach was around 10.1 billion US dollars. Moreover, in the year it had cost around 9.23 billion US dollars. If it is studied carefully it can be analyzed that there has been an increase in the cost. From the year 2020 the pandemic has started and data security has become a grave concern for all of the country. Data mining technology can be used in the health sector to improve the data breach and the useful data regarding the drugs and medicines will be safe.

According to the graph the data breach in the pharmacy sector was also significant and this rise is due to research and innovation for the developing the vaccines against the covid. The total cost for the data breach for the pharmacy sector was

around 5.01 billion US dollars. Moreover, the cost in the technology that can be used for the data breach is around 4.97 billion US dollars.

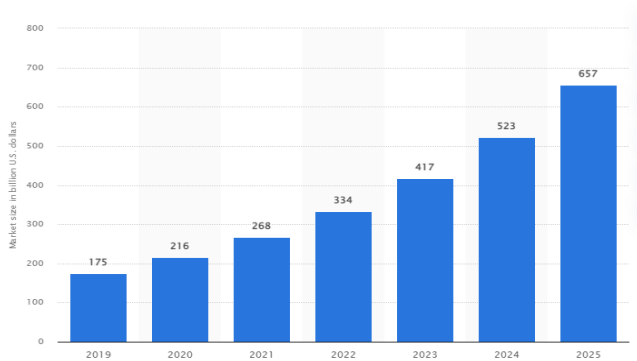


**Figure 1:** Global average data breach cost 2020-2022 by industry

With the help of data mining it is very much possible that malware can be detected so that the threats of the cyber attacks will be minimized. Many a time the fraud detection is not possible these are also using the latest pattern and they try to come with those patterns. If the system software fails to understand the pattern then they can be considered as the data theft. There are drawbacks of this process as they have a limited number of patterns and if any of the new patterns appears it fails to explain and further it has generally made the software vulnerable [8]. If data mining is included in the healthcare sector it has a lot of benefits like improved clinical decisions, better management of the customer with the company. In the treatment of the people, precision for precise diagnosing of the disease. Data mining technology can also be used in the hospital administration as which patient needs which type of care and other things. Many a time, an unknown disease appears in front of doctors that are not understood by them. This frustrates the healthcare professional but if they use the data mining technology it is very much possible that the trends of that disease can be predicted easily.

As digitization is going across the world, the trends of digitization in the health market are also increasing. The graph that has been provided is based on the global healthcare market on a digital basis. According to the graph that has been provided by the year 2025 the total market size of the digital market would be around 657 billion US dollars. The prediction has been started from the year 2019 when the total market size was estimated to 175 billion US dollars. As the

market size of the digital market is going to increase in the near future it can be said that the use of data mining is going to increase. Data mining process has some of the drawbacks and those drawbacks are as follows; data mining is a complex art and thus it is difficult to become an expert of the process. The techniques are very useful but at the same time they are not infallible in nature [10]. Data mining raises privacy concerns and they are quite expensive in nature and hence is not applied by all of the organization. Instead of these drawbacks it can be said that data mining is still the best technology for data analytics.

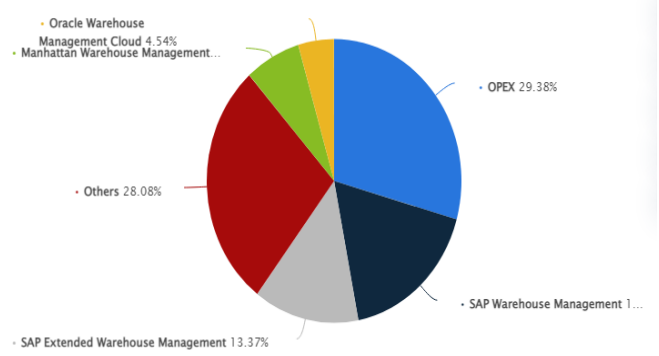


**Figure 2:** Global digital health market

**OLAP and its uses**

OLAP or Online Analytical Processing can be defined as a set of approaches that is useful in representing data from multiple dimensions. It basically includes a bunch of practices that aim at modeling and database and further creating specific solutions based on analytics. There are three main types of OLAP that are relational, multidimensional and hybrid [11]. *Relational OLAP*, in relational OLAP the data is stored in a relational database as it handles a large amount of data. *Multidimensional OLAP* is the type in which the data on the dishes are in a specialized manner called the multidimensional array. *Hybrid OLAP* is the type of the where the data is stored in large amounts and it is further detailed data. There is wide range of uses of OLAP and some of its uses are data mining, complex calculations and business reporting. It can also be used in budgeting and forecasting along with financial analysis.

The application of OLAP could be understood with the help of the given example like any type that is used in the warehouse management can be said that it is based on the OLAP system. Companies like spottily use the technology to make their homepage according to the likes and dislikes of its customers. The feature personalize homepage is the feature that is not possible to provide by the company to the user without OLAP. Another example is the features that are recommended by the applications to the users. Like YouTube, Netflix, Amazon prime are some of the examples of the companies which often provide recommended content for them according to their interest. This feature can only be possible with the technology of OLAP.



**Figure 3:** Warehouse management software market world share by the year 2022

The above graph has shown the vendors that have made the most impact on the warehouse management. Oracle warehouse management has the minimal share globally and the lion share is gained by the OPEX. In the warehouse management the thing which involves in the process is the receiving, tracking, and storing. Along with this training of staff, managing shipping and continuous monitoring of goods are the tasks which are performed in the warehouse management [13]. The technology of OLAP helps to ease the process with the help of technologies like data mining. The technology of data mining and big data analytics can be used for making the process more and more convenient when it comes to managing a large amount of data. Moreover, there is a process called that that can be seen in the process and that can be predicted as well.

Many of the warehouse companies had used the technology of OLAP in the area house management and they had achieved significant growth. Globally, Oracle warehouse Management Company is considered the best warehouse management company and that is due to its best and quality service using the OLAP technology. These are some of the positives and now the negatives have been discussed here like:-

Restructuring of data is often required by the OLAP and could not sustain long if the data is not provided in proper format. In OLAP there are a limited number of dimensions and this is also considered as the drawback of OLAP. It is often said that the process of changing the OLAP cube to be fully updated is often lengthy. These are some of the drawbacks of the OLAP but now the importance of OLAP for the warehouse management will be discussed. In the warehouse management there are five benefits of use the OLAP in the warehouse management. Those benefits are business focused multidimensional data, business focused calculation, and data calculation is trustworthy in nature, flexibility in the process and speed of thought analysis [14]. The technological progress has shown that in future the technology is progressing and the OLAP technology can be used in the healthcare sector unlike the warehouse management. The main competitors of OLAP can be considered as tableau, micro strategy and congas.

### **Difference between the OLAP and Data mining**

Both the data mining and OLAP are different and that too in a sense of using the technology for the process of utilizing the data as much as possible. The approach that is followed by data mining is the bottom up approach whereas the approach that is followed by OLAP is the top down approach. The top down can be understood as the approach of management in which the key decisions are taken by the top sources. Whereas, the bottom approach is the approach in which starts from the grass root level and it reaches to the top management position. This can be said as the vice-versa of the top down approach. The data mining process involves the in-depth analysis of data with the help available input data [15]. This helps an organization in making better decisions and giving better predictions. Therefore for future perspective data mining is a useful tool. But at the same time the tool OLAP is used to analyze the data of the past. The data mining approach is the type of approach that is driven by discovery whereas the OLAP is driven by query. In discovering a new possibility in any of the processes comes the technological tool data mining. But in between the process if any of the queries appears the analysis will be done at that time the tool that will be used is OLAP.

The technology of data mining is becoming popular data by day and it is not used widely by most of the companies as it needs good expertise and it is overall a costly approach. At the same time the data that has the OLAP tool is well aware to the people and it is not an emerging tool and they are used widely by many of the companies. The main motto behind using the technological tool data mining is to develop an insight and predict what is going to happen in future [16]. On the other hand the OLAP approach is used to make a better analysis of the given but not to predict any of the possible future outcomes. Data mining often focuses mainly on the data that are provided in the format of summary and the OLAP deals with the detailed and in depth transaction level data. The main focus of data mining is the large amount of data whereas the OLAP supports complex calculations and in this way it can be said that the process of making data mining tools is different from the OLAP.

Companies like MC Donald's, Netflix, Amazon and various other companies actively use the technology of making the process more and more convenient and predict the like and taste of the customer. The companies like Amazon, ASG software, Gray matter software and many more are the companies that use OLAP technology for the betterment of their business process [17]. It can also be said that the data mining process is a subsidiary of the OLAP as for using the OLAP of the good sort of data the technology that is used is the data analytics technology. In the given figure it has been given that the OLAP technology is mostly used by the warehouse management company but the technology can also be used by the other sector as well where complex calculation is required. Also the data mining can be used in the medical and healthcare sector as that will be remarkable for making the process more and more lucid and

simple. Data analytics technology is an emerging technology in the future as it is costly and the technology needs a good sort of knowledge of data science but in the future it is estimated that it will be eased further [18]. Therefore if the data analytics technology will be mastered in future it can be said that this is going to be a game changer for the organization.

### **DISCUSSION**

Data mining technology is energy technology of data analysis which needs good knowledge of data science and statistics. Data mining process was coined many years ago but most of its impact can be understood now as it has made progress in the field of health care, marketing, influencing people for the purpose and many more. In this way it can be said that in the future the process of data mining will be a game changer in future. The technology of OLAP can be understood as the technological tool that is used to answer the queries with a multidimensional approach and that too swiftly in computing. In the result section it is trying to provide a brief account of the difference between data mining and OLAP. Here the difference has been discussed with first explaining both the topic and then its difference. The graph that has been provided in the section is about the cost of data breach in the various sectors like the health sector, financial, pharmacy, energy and public enterprises. After understanding the graph it can be understood how significant the use of data privacy or the data security is.

Therefore for that purpose the data mining will be the most effective tool as it understands the pattern of the breach and thus in future if any such breach appears it can easily be minimized. The next graph that has been provided here in the results section is about the global market digital healthcare where it had predicted by statista that by the year 2025 the total size of the digital healthcare market will be around 657 billion us dollars. If the size of the healthcare market will increase further in future it can be said that the use of data mining methods will be crucial. The infrastructure of digital healthcare started from the year 2019 just before the pandemic and after the pandemic the investment in the digital healthcare had been increased. The third figure that has been provided here is pie chart and this is where the different warehouse management companies have been dominating the global markets and due to this it can be understood that warehouse facilities are also using the technology like data mining and OLAP. With that it can be understood that the warehouse today is using some of the technology and the technology using it OLAP is the most useful tool for the warehouse management companies. The share of the Oracle warehouse management company is the least in the graph but when it comes to the quality service it is the best warehouse management company. Hence, it can be said that both the technology OLAP and data mining has its own significance but for the future prediction the data mining is better as compared to the OLAP.

## CONCLUSION

In the following study, the subject matter has been surrounded by the developing conception about differences between OLAP and data mining and this study has been served with authentic and thoughtful insights which are related with the subject matter. At the beginning of the study, the concept of OLAP and data mining has been introduced and a minor difference has been developed between two following procedures. The data mining and OLAP have been served at the beginning of the study with proper elaboration in an extended. After the elaboration of the concept of data mining and OLAP, the material and methods have been selected for the following study. For the study, the 4 inductive approach, cross-sectional research design and secondary data has been selected for this given study. Later on the topic of discussion has been served with proper sorts of themes which depict the differences between data mining and OLAP in a certain manner. Also, the drawbacks of using data mining have been represented in a certain manner.

The concept of data mining has been depicted and the use of the procedure of data mining has been represented in this following article on the health sectors and hospitals. The OLAP has served as the core process of multiple data transcription which has been related with the following subject matter of the study. Also, the following study has been served with valid conclusions which have been maintained with proper insights and the study has been executed within proper manner by gathering valid data.

## REFERENCES

- [1] Shelest, Oleksandr, and Bella Holub. "Olap Systems As The Modern Data Preparation Tools For Outdoor Advertising Data Mining." *Three Seas Economic Journal* 1.2 (2020): 60-66.
- [2] Li, Xiaodong, et al. "What are the occupational hazards of construction project managers: A data mining analysis in China." *Safety science* 134 (2021): 105088.
- [3] Neziraj, EminQerim, and AferditaBerishaShaqiri. "The Business Process Needs for Excellent OLAP and Data Mining Tools in Kosovo Companies." *EuroEconomica* 38.1 (2019).
- [4] Braun, Virginia, and Victoria Clarke. "Reflecting on reflexive thematic analysis." *Qualitative research in sport, exercise and health* 11.4 (2019): 589-597.
- [5] Santos-Pereira, Judith, Le Gruenwald, and Jorge Bernardino. "Top data mining tools for the healthcare industry." *Journal of King Saud University-Computer and Information Sciences* 34.8 (2022): 4968-4982.
- [6] Tomasevic, Nikola, Nikola Gvozdenovic, and Sanja Vranes. "An overview and comparison of supervised data mining techniques for student exam performance prediction." *Computers & education* 143 (2020): 103676.
- [7] Statista Research department. "Average total cost per data breach worldwide 2020-2022, by industry." Statista, 6 September 2022, <https://www.statista.com/statistics/387861/cost-data-breach-by-industry/> Accessed 10 January 2023.
- [8] Hanif, Hazim, et al. "The rise of software vulnerability: Taxonomy of software vulnerabilities detection and machine learning approaches." *Journal of Network and Computer Applications* 179 (2021): 103009.
- [9] Stewart, Conor. "Global big data healthcare analytics market size by application 2016 & 2025." Statista, 19 October. 2022, <https://www.statista.com/statistics/909669/global-big-data-in-healthcare-analytics-market-size-by-application/>
- [10] Tomasevic, Nikola, Nikola Gvozdenovic, and Sanja Vranes. "An overview and comparison of supervised data mining techniques for student exam performance prediction." *Computers & education* 143 (2020): 103676.
- [11] Patel, Jigna Ashish, and Priyanka Sharma. "Online analytical processing for business intelligence in big data." *Big Data* 8.6 (2020): 501-518.
- [12] Vailshery, Lionel Sujay. "Share of global warehouse management software market by vendor 2022." Statista, 13 September. 2022, <https://www.statista.com/statistics/503241/worldwide-data-warehouse-management-software-market-share/>.
- [13] Queiroz-Sousa, Paulo Orlando, and Ana Carolina Salgado. "A review on OLAP technologies applied to information networks." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 14.1 (2019): 1-25.
- [14] Lamani, Asma, et al. "Data mining techniques application for prediction in OLAP cube." *International Journal of Electrical & Computer Engineering* (2088-8708) 9.3 (2019).
- [15] Ben Kraiem, Maha, et al. "OLAP operators for social network analysis." *Cluster Computing* 23.3 (2020): 2347-2374.
- [16] Yang, Jin, et al. "Brief introduction of medical database and data mining technology in big data era." *Journal of Evidence-Based Medicine* 13.1 (2020): 57-69.
- [17] WM Ribeiro, Marcello, Alexandre AB Lima, and Daniel de Oliveira. "OLAP parallel query processing in clouds with C-ParGRES." *Concurrency and Computation: Practice and Experience* 32.7 (2020): e5590.
- [18] ur Rehman, Muhammad Habib, et al. "The role of big data analytics in industrial Internet of Things." *Future Generation Computer Systems* 99 (2019): 247-259.