**ADVANCED DATA MINING TECHNIQUES**

**Dr Ani Brown Mary N, Dr Stella Janci Rani J, Dr Adline Rajasenah Merryton, Dr Richlin Selina Jebakumari A**

**Assistant Professor, Department of Computer Science**

**Sarah Tucker college**

**Corresponding Author emailID:** [**anibrownvimal@gmail.com**](mailto:anibrownvimal@gmail.com)

**I INTRODUCTION**

Data mining is an essential component of knowledge discovery in computer science that tries to extract useful patterns and insights from massive databases. Data mining techniques are used to evaluate and turn raw data into information that can be utilized to improve products and services, make better business decisions, and improve consumer experiences. Numerous fields employ data mining extensively. They are Healthcare, Education, Telecommunication, Classification, Criminal Investigation and Intrusion Detection. These tools help businesses streamline processes, enhance customer satisfaction, lower risks, and make data-driven decisions. Despite the fact that there are many commercial data mining technologies accessible today, there are still several difficulties in this area.

Classification process analysis can be used to find important and pertinent data information, as well as metadata. This method of data mining helps to categorize data into various categories. The clustering process is a mining technique that helps to find data that are similar to each other. This helps in understanding the similarities and differences between the data. Regression analysis is a method for mining data that identifies and studies the relationships between variables. It can be used to determine the probabilities of a certain variable in light of the presence of additional variables. Association Rules method of data mining assists to discover the relationship between two and more Items. It finds an undiscovered pattern within the set of data.

Outer detection is a particular type of data mining method is the observations of data elements in the data that don't match the pattern expected or behavior. This technique is used in a wide range of areas like intrusion detection of fraud, fault detection, and so on. Outer detection can also be referred to as Outlier Analysis and Outlier mining. Sequential patterns is a technique of data mining helps to identify or discover patterns or trends that are similar in transactions for a specific period. Mining prediction is a mix with other methods of data mining such as patterns that repeat, trends and clustering, classification, and more. Studies past events or events in a proper sequence to anticipate the outcome of a future event. Hence, it is an important component of the KDD process in data mining. The tracking of patterns is an essential method for data mining. It involves identifying and monitoring patterns or trends that are present in the data to draw predictions about business results. When an organization has identified trends for sales statistics, say it's time to consider making a decision to take advantage of that knowledge.

**II MARKET BASKET ANALYSIS**

Market basket analysis is a data mining approach that merchants employ to improve revenue by better understanding client buying habits. Large data sets, such purchase histories, are analyzed to identify product groups and those that are more likely to be bought together. Retailers can utilize this technique to discover more about the preferences of their customers. Retailers can use this information to change the layout of their stores, which will simplify and speed up client shopping. The retailer may make necessary layout changes to the store based on this information to better understand client needs. It is possible to compare client data from numerous stores and from various demographic groups using different analytical techniques. Because it can assist merchants in identifying customer behavior patterns and creating customized marketing strategies, data mining is crucial to Market Basket Analysis (MBA). Identifying cross-selling opportunities, anticipating consumer behavior, and optimizing pricing methods are a few popular applications. For instance, grocery companies utilize data mining to evaluate client purchasing patterns and find relationships between products. Pattern recognition and the prediction of probable defections provided by data mining enable customer retention. The Risk Assessment and Fraud departments also employ the data-mining methodology to spot erroneous or unusual activities.

The goal of market basket analysis is to discover the goods or groups of products that frequently appear in purchasing transactions. The information gleaned from a market basket analysis can be quite beneficial; for instance, a supermarket could use it to restructure its layout by grouping together items that are commonly purchased together and placing them close together. However, it can also be used to increase the effectiveness of a marketing campaign: related products shouldn't be promoted at the same time. It should be possible to boost sales of only one of the related products while generating parallel sales growth for the other related products [1]. Xie et al [2] propose Market Basket Analysis Based on Text Segmentation and Association Rule Mining. Market basket analysis, which looks for associations between the products people collectively buy, is very useful in assisting trade markets with making scientific decisions. Maintaining a strong position in the market is always difficult for organizations in a world of competitive markets since it always depends on those companies' capacity for decision-making and comprehension of consumer behavior [3]. Market Basket Analysis, also known as association rule learning or affinity analysis, is essentially a data mining technique that is extensively utilized in the fields of marketing, nuclear science, education, and bioinformatics [4]. The Association Rule is the primary tool used in MBA [5] i.e., the IF, THEN construct. IF a client purchases bread, THEN he is expected to purchase Jam as well. These enable the prediction of client behavior trends. As a result, it is possible to put together specific offers with products that buyers are likely to purchase. As a result, the company's sales and revenue will rise automatically [6].

**III VARIOUS MARKET BASKET ANALYSIS METHODS**

Market Basket Analysis comes in three different Flavors. They are Descriptive, Predictive and Differential Market Basket Analysis.

**A Descriptive MBA**

The main purpose of this form of research is to comprehend customer behavior, including the most frequent item combinations and the products that are purchased in combinations. By using descriptive market basket analysis, retailers can more profitably put products in their stores by learning which products are frequently purchased together.

**B Predictive MBA**

Market basket analysis that predicts future purchases based on historical buying patterns is known as predictive market basket analysis. In order to make predictions about which items are most likely to be purchased together in the future, large volumes of data are evaluated using machine learning algorithms. Retailers can utilize predictive market basket [6] research to make data-driven decisions regarding what products to stock, how to price them, and how to optimize shop layouts.

**C Differential MBA**

Analyzing competitors is benefited by this kind of analysis. To discover intriguing trends in consumer behavior, it compares purchase histories across outlets, seasons, two time periods, different days of the week, etc. [5]. With the aid of differential market basket analysis, retailers may adapt their marketing and sales strategies in response to changing consumer behavior.

**III SOCIAL MEDIA OPTIMIZATION**

SMO enables us to create a social network and engage with others as well as your consumers. We can also read other people's tweets and status updates to learn about their interests, preferences, and expectations. As we modify social media advertising, it is possible for us to target and retarget. For example, we may target individuals based on their region, education level, or even their purchase history and pages they have liked. It empowers us to reply swiftly to client inquiries and complaints regarding our product or service. We can immediately give appropriate information and handle problems to please our consumers. It enhances brand loyalty; for example, companies that are engaged on social media have more devoted consumers.

The use of social media sites to manage and improve an organization's messaging and online presence is known as Social Media Optimization (SMO). It may be used as a digital marketing approach to raise awareness of new products and services, engage with consumers, and neutralize potentially bad news. It is also used to improve a company’s or organization’s brand and online presence. Social media platforms have been used to raise awareness of new products and services and to assist businesses in connecting with customers. It enables us to deliver targeted visitors with social network updates and to create a social network and communicate with people and customers. We may also learn about other people's hobbies, preferences, and expectations by reading their tweets and status updates.

Further, SMO gives us a competitive advantage. We can approach potential clients before your competition. Today, more than 90% of companies use multiple social media platforms. Moreover, it assists us in increasing your sales through social media marketing. It is estimated that over 70% of business-to-consumer marketers have recruited consumers using Facebook. It broadens our consumer base, allowing the customers whom we didn’t know to locate us easily.

SMO is a method or collection of actions that aids in the rise of website traffic using social media sites such as Facebook, Twitter, LinkedIn, and Google+. A social media site is an online platform that allows us to engage with individuals from all over the world and develop social networks. We can use social media to do a variety of things, such as engage in debates, express your opinions, build a page to advertise your business, and so on. Social media are digital communication platforms that makes it possible to share text and multimedia content through online networks and communities. Social Media facilitate the conception and sharing of information, ideas, interests, and other forms of expression through virtual communities and networks. Social media groups can provide:

* a sense of belonging and community.
* access to support from like-minded individuals.
* an appreciation of different perspectives.
* reduced isolation.
* a monitored discussion environment.

Making anything like a design, system, or choice as entirely flawless, functional, or effective as is humanly feasible is known as optimization. specifically: the mathematical techniques used in this, such as determining the maximum of a function. Steps of Optimization are as follows:

* Define: The first step to is define the processes that require optimization.
* Measure: The second step is to measure and identify how the process performs.
* Analyze: The third step is to analyze how you can optimize the process.
* Improve: The fourth step is to improve the process.

**A Social Media Optimization Instructions**

Identify Your Social Media Goals: The first part of the process is to identify the ultimate goal for building and optimizing social media accounts. For instance, if our goal is to utilize the social media accounts to boost sales, we have to take necessary steps to increase it.

Fill Out the Profiles fully: when optimizing your social media pages is to fill out your main profile pages front to back stating what our page is all about. In order to reach our target audience, we have to make your profile visible in search results.

Use Hashtags: Hashtags are essential in social media platforms like Twitter and Instagram. Adding hashtags to the captions increases our chances of exposure. Adding one or two hashtags per tweet is best practice. In fact, studies show that tweets with only one hashtag generate the most engagement.

Create Valuable Content That Solves Your Followers' Problems: Content is the key to our social media marketing success. People love things that provide value to them. Thus, our content, whether created or curated, should not be limited to plain promotional materials.

Use Images: The right images can make a person notice us. This is something all brands need in order to increase engagement, especially with the enormous amount of content lining up in every user’s feed, fighting for attention.

Use Attention-Grabbing Headlines: Headlines matter and can make or break it when it comes to getting our followers to stop for a second look. Using negative words like “worst” and “never” can sometimes have better results compared to positive words.

Track the Progress: The only way to see if our social media accounts are properly optimized is by tracking the progress. See how much engagement we’re getting. Take note of the number of times our post was shared, the number of comments each post received, the number of likes, etc.

**IV DATA MINING APPLICATIONS IN RESEARCH ANALYSIS**

Here are some of the most popular industries where data mining is used. The diagram of datamining applications is shown in the Figure 1.

****

**Figure 1 Data Mining Applications**

**A Data Mining in Healthcare**

Data mining is becoming more and more important in the healthcare sector as a result of the massive amount of patient data that is created daily. Healthcare firms can lower costs and improve patient outcomes by identifying best practices based on data and analytics. Multidimensional databases, machine learning, soft computing, data visualization, and statistics are some of the data mining techniques employed by researchers. It can be useful for predicting patients from various categories. Data mining can be used by healthcare insurers to spot fraud and abuse. In general, these can be categorized as the assessment of treatment efficacy, healthcare management, customer relationship management, and fraud and abuse detection [7]. Finding effective standardized treatments for particular diseases can be aided by data mining. Medical research employs data mining to examine patient medical information and pinpoint elements that contribute to the development of diseases. Data mining is mostly utilized in the healthcare industry to diagnose patients and support doctors' clinical decisions by helping to forecast certain diseases [8]. Healthcare includes intricate procedures for the identification, mitigation, and prevention of illness, injury, and other physical and mental disabilities in people [9].

**B Data Mining in Education**

The 21st century society is continually changing as a result of the rapid and ongoing advancement of technology. One setting where these technology advancements are extremely applicable and where adjustments to teaching methods and resources are required is the educational setting. The manner in which the teaching-learning process is carried out is crucial, especially in higher education [10]. This process, whether it is F2F teaching, blended learning, or e-learning, is occurring more frequently within LMSs. The goals of EDM are acknowledged to include advancing learning science, researching the effects of educational support, and endorsing students' prospective learning behaviours. Teachers require tools to assist them in their educational activity, which must go beyond only dispensing knowledge, as well as in keeping track of their students' learning progress. The development of teacher preparation in active pedagogies, the incorporation of student tracking modules with artificial intelligence and data mining, and LMS tools. The use of blended learning and e-learning in LMSs (like Moodle) for teaching in higher education is on the rise.

**C Data mining in Telecommunication**

Data mining is used by telecommunications businesses to enhance marketing campaigns, spot fraud, and maintain their networks more effectively. Due to the abundance of high-quality data accessible, the industry's competitiveness, and the developments in data mining, data mining should play a significant and growing role in the telecommunications sector [11]. In the telecommunications sector, data mining aids in pattern recognition, fraud detection [12], efficient resource management, and service quality enhancement. Important industrial firms may be able to enhance the quality of their services through data mining to remain competitive. Pattern analysis of spatiotemporal databases can considerably enhance mobile communications, mobile computing, as well as online and information services. Furthermore, methods like outlier analysis can be used to find fraudulent users. Additionally, Online Analytical Processing (OLAP) and visualization tools can be used to compare data on user group behavior, revenue, data traffic, system overloads, and other topics.

**D Data mining in Classification**

One of the significant techniques in data mining is classification. This technique is applied broadly across many industries. The classification technique in data mining aids in classifying data into distinct groups in order to obtain important metadata. The objective of classification in data mining is to assign a class label to each instance in a collection based on its attributes. Making a model that reliably predicts the class labels of upcoming instances based on their features is the aim of classification. These days, it is employed more and more in research and technology to extract the enormous volume of data. Naive Bayes Classifier, Decision Tree, Neural Networks, and Support Vector Machine are a few examples of classification techniques [13]. Data mining techniques like classification are frequently utilized in areas like sentiment analysis, email screening, and medical diagnosis. For example, Techniques for classification make it easier to separate important criteria from unimportant ones that affect consumers' banking decision. Additionally, the identification of consumers with comparable loan payment behaviors is made possible using multidimensional clustering approaches. The laundering of money and other financial offenses can be found with the aid of data analysis and mining.

**E Data mining in Criminal Investigation**

Data mining is acknowledged as a relatively recent and in-demand field of study as a tool for criminal analysis. Crime analysis, which is a large category of research disciplines, can include a wide range of criminal behaviors, from minor infractions of the law to globally organized crime [14]. This is understandable given that data mining is a relatively new and rapidly developing field. Those interested in historical and contemporary definitions of data mining are directed to [15] as the goal of data mining is to discover models through an algorithmic search process that explores linear and nonlinear models, whether explicit or not. Criminal profiling, study of criminal networks, and prediction of crime hotspots are popular data mining applications in the field of criminal justice. Analysis of crime data is used to identify crime hotspots, or locations where crimes are more likely to occur. Criminal profiling entails looking at crime statistics as well as demographic information to find prospective culprits. Criminal network analysis examines social network data to find links between offenders. Criminology is a perfect field for applying data mining techniques due to the vast amount of crime datasets and the intricate interactions between them. Crime reports written in text format can be converted into word processing files. These facts can be used in a process for matching crimes.

**F Data mining in Intrusion Detection**

Identification of potential security risks to computer networks and systems is a component of intrusion detection. Data mining is crucial to intrusion detection since it can be used to spot patterns of harmful behavior and create security plans. Network intrusion detection, malware detection, and spam filtering are some popular uses. Data mining, for instance, is employed in network security to examine network traffic and spot potential security risks. Increased internet usage and the availability of tools and techniques for hacking and attacking networks led to the development of intrusion detection as a crucial aspect of network administration. Therefore, intrusion detection is required as an additional barrier to safeguard computer systems [16]. The core components of intrusion detection are the following: resources in a target system that need to be protected, such as user accounts, file systems, system kernels, etc.

**References**

1. Trnka, A. (2010). Market Basket Analysis with Data Mining methods.
2. Xie Wen-Xiu, Qi Heng-nian and Huang Mei-li, "Market Basket Analysis Based on Text Segmentation and Association Rule Mining", First International Conference on Networking and Distributed Computing (ICNDC), Pp. 309 – 313, 2010.
3. Raorane, A.A., Kulkarni, R.V. and Jitkar, B.D., 2012. Association rule–extracting knowledge using market basket analysis. Research Journal of Recent Sciences ISSN, 2277, p.2502.
4. Dubey, S. K., Mittal, S., Chattani, S., & Shukla, V. K. (2021). Comparative Analysis of Market Basket Analysis through Data Mining Techniques. 2021
5. https://www.javatpoint.com/market-basket-analysis-in-data-mining
6. <https://www.geeksforgeeks.org/market-basket-analysis-in-data-mining/>
7. Hian Chye Koh and Gerald Tan, “Data Mining Applications in Healthcare”.
8. Neesha Jothi, Nur Aini Abdul Rashid, Wahidah Husain (2015), Data Mining in Healthcare – A Review
9. J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, and H. Pan, “Emerging information technologies for enhanced healthcare,” Comput. Ind., vol. 69, pp. 3–11, 2015.
10. María Consuelo Sáiz-Manzanares, Díez, José, Juan José Rodríguez-Díez,, Raúl Marticorena-Sánchez and Yi Peng Ji “Monitoring of Student Learning in Learning Management Systems: An Application of Educational Data Mining Techniques”,2021.
11. Gary M. Weiss, Data Mining in the Telecommunications Industry,2009
12. Mohsin Nadaf & Vidya Kadam (2013),” Data Mining in Telecommunication”.
13. S. Umadevi; K. S. Jeen Marseline, “A survey on data mining classification algorithms”,2017
14. H. Chen, W. Chung, Y. Qin, M. Chau, J. J. Xu, G. Wang, R. Zheng, and H. Atabakhsh, Crime data mining: an overview and case studies, In Proceedings of the 2003 Annual National Conference on Digital Government Research, Boston, MA, 2003, 1–5.
15. H. Hassani, G. Saporta, and E. S. Silva, Data mining and official statistics: the past, the present and the future, Big Data 2 (1) (2014), 34–43.
16. Wenke Lee Salvatore J. Stolfo.” Data Mining Approaches for Intrusion Detection”