Technoarete Transactions on Intelligent Data Mining and Knowledge Discovery
Vol-2, Issue-1, February 2022
e-ISSN: 2583-195X

Necessity of Data Mining for Information Security

Dr. V. Gokula Krishnan[1]*, Dr. Ajay B Gadicha[2]

[1] Professor, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Thandalam, Chennai, Tamil Nadu, India.
[2] P R Pote Patil College of Engineering and Management Amravati, India.
*Corresponding Author Email: [1] gokul_kris143@yahoo.com

Abstract

Data mining is the technology that has been used in wide areas some of that area like the analysis of data for giving the best possible products for the customer by the ecommerce companies. There are many other areas where the data mining techniques can be used in information security. This discussion has given an account of what are the necessity of using the data mining for the information security. It had been explained in four different sections namely introduction where a brief background about the research has been provided. Later on the methodology section which provides the information about the type od research that has been used for the research. After this a results section have been discussed here an in-depth analysis has been done on the research topic. The results section has aimed to provide a detailed rationale of the topic with the help of data and graphs. After this part a discussion part have been taken here a brief discussion have been provided regarding results part. Finally, this research study has been concluded with the help of the conclusion part where each of the useful findings of each section have been discussed.

Keywords

Data mining, Security, Technology.

INTRODUCTION

Data mining can be explained as the process of uncovering the pattern and important information from a large set of data. In other words it can be said that for improving the market segmentation and exploring the large database it can only be possible with the help of data mining or data analytics [1]. E-commerce platforms like Amazon and Alibaba have used this technology so that the taste of the customer and like and dislike of the customers can be analyzed thoroughly. For getting a competitive edge in the highly competitive business environment it has become important to make the best possible analysis. These companies often use the data of the people from the various platforms as well as their own platform which comes with the features like the filters. Those filters can be applied on the basis of the different parameters that have been set by the company itself. Those parameters are like taste of the customer, gender of the customer, and like & dislike of the customers [2]. Those filters actually help to minimize those products that are often meaningless for a particular customer and give them the best possible products according to their choice.

There are various types of data mining and those are like pictorial mining, web mining, and audio mining. Hence it can be said that with the help of data mining and data analysis it is possible to make better marketing decisions and palace the products in the right places. Information security can be explained as protecting the sensitive and confidential information for the unauthorized persons in the market which include the inspection, modification and redesigning. The role of data security or information security is becoming important these days as more and more companies are providing their information online and thus data theft is the most possible threat [3]. In this research it has been taken very seriously and thus this research study is going to provide the necessity of data mining for information security.

MATERIALS AND METHODS

The section of this study is going to discuss the material and methodology that have been used for the conduction the research which can be said to be the research methodology. The research methodology that has been discussed here is the data collection method aligned with the process of analyzing the data [4]. In this research the data had been collected on the basis of a secondary data collection method. The collected data had been further analyzed with the help of the thematic data analysis method. The research topic is about the necessity of data mining for information security hence it needs to discuss the importance of data mining these days. Importance of information security for an organization and after that in what ways data mining can be a game changer in the field of energy and data security [5]. These are some of the things which are necessary to be discussed with the recent developments and for that purpose thematic analysis has been done to understand each and every theme in depth. A thematic analysis generally includes studying the journals, books and research papers that have been published by the other researcher. For this research the data have been taken on secondary basis and further analysis with the help of thematic data analysis.

RESULTS

Data mining is done these days for various purposes and among them the most significant steps are risk management, mitigating cyber security threats, and other complex business uses. The online users have been increasing day by day globally and the digital footprint is also increasing [6]. Therefore, for the betterment of the functioning a company

e-ISSN: 2583-195X

needs to make use of the data mining process. The market of the data analysis and the revenue that is generated from their operation is huge and this could be understood with the help of a health that have been given below.

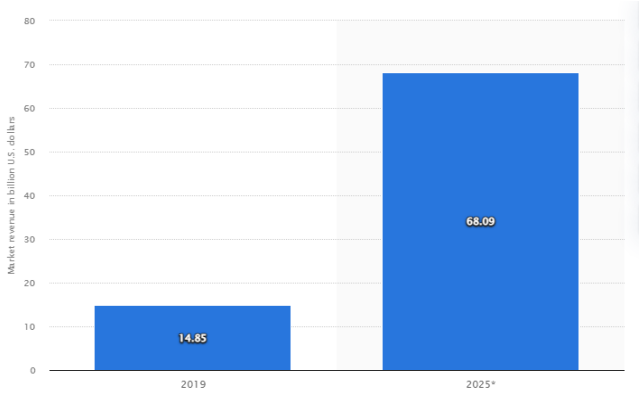

Figure 1: Data Analytics market size in the year 2019 and prediction of the year 2025

Globally the data market size has been increasing at a significant pace and it has grown at the speed of compound growth structure. According to statista, the total market size for data analysts in the year 2019 was around 14.85 billion dollars and it seems to rise by almost 30% in the coming years [7]. If the growth rate remains the same as it is now it can be estimated by the year 2025 the overall market size of data analytics to become 68.09 billion US dollars. Technological innovations are going across the world at a rapid pace and many of the companies are uploading to this process of collecting and analyzing the data. In the future it can be estimated that the revenue with the process of data mining seems to rise in the coming years. During the time of pandemic most of the business operations were shut but the technology was making high incomes and in the time of pandemic the demand of the technology did not reduce.

Technologically advanced countries like the USA, UK, Australia, and Japan use these technologies for the development of their businesses. It can be said the term data mining is similar to data analysis, Artificial intelligence, and machine learning. These days' data can be created with the help of social media and online blogging. The other benefits are like improving the data storage by entities like social media companies. The best examples of the social media companies are Face book, Twitter, YouTube and Whatsapp others use the technology of data mining.

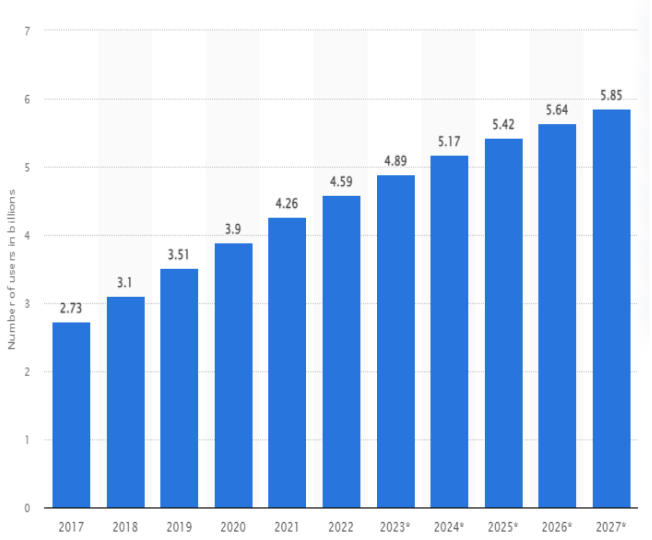

Figure 2: Global Social Network users

The figure above has given a brief description of how many users of social media there are globally. According to the graph that has been given in the year 2017 the total number of social media users was around 2.72 billion users. In the next year the figure went up and now the total number of users has become 3.1 billion users [8]. From the year 2017 to the year 2027 total number of social media users seems to rise exponentially. As of the year 2022 the total number of social media users is around 4.59 billion dollars. According to the united nation the total global population will reach the 8 billion mark [9]. It has also been estimated that the people in India and China will be the highest in the world and India seems to surpass China in the coming years. The market of both the countries is huge and the Smartphone users have also increased in recent years in India and China.

The statistics that have been shared in the above figure is a clear cut estimate that the share of the social network users will be dominated by the Indian market as in China the Social media platforms like face book, YouTube, twitter and others are not fully operational. According to Statista, if these trends follow for long then in the year 2023 the total number of active social media users will be around 4.89 billion users. That is way more than the previous year that the year 2022. Moreover, if the number of social media users rises in the coming year the online threats seem to increase. It could be in the form of cyber security threat and data theft [10]. In the coming years if social media users seem to increase in the coming years then data mining will be the most effective and significant weapon for data security.

Product based companies and commercial food service companies like McDonald's use the data of the customer form the Social media platform to understand the trends as well as the current likes and dislikes of the company. According to the like and dislike of the people as well as the customer many of the companies make their marketing of the product strategy. It can only be done by the careful study of the data of the people and that can be done with the help of data analysis. Including the data mining in the information

e-ISSN: 2583-195X

security has many benefits like it helps to find the security flaw and blind spots in the process also it helps in detection of the zero-day attacks. It also has some of the cons like the whole process needs good expertise or the person has to be an expert in data science. The process of data mining is time taking and it needs a huge effort from the data scientists. Data mining is often checked manually and it is also done by manually there is a constant threat of disclosing sensitive information [11]. Data mining can be applied with the different type of techniques that are involved and as of now there are six different techniques that are often followed for implying the data mining.

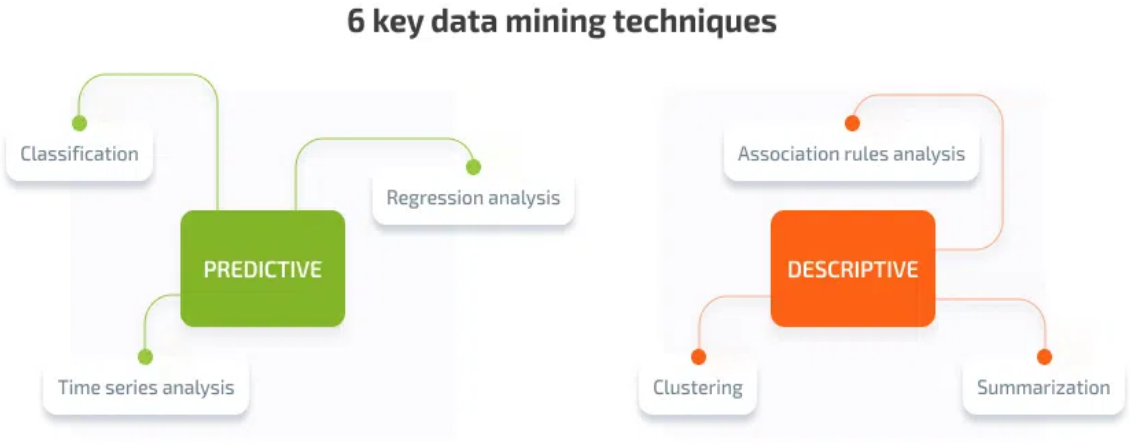

Figure 3: Different types of Data mining techniques

The six different steps by which the data mining can be performed is the Classification, regression analysis, time series analysis, clustering, summarisation, and Association rules analysis. These are the techniques that are used for the data mining for critical analysis and understanding the market trends. Due to the availability of the huge amount of data in the online platform it is very much possible that in future all information security will be the most significant issue. For the effective cyber security data mining process can be used like detection of malware, detection of intrusion, detection of fraud, threat intelligence and insider threat detection as well as prediction. ***Malware detection,*** many times when software is used for a long time some of malware appears in the software and it needs to be removed [12]. It generally reduces the speed and affects the performance of the application software for the proper detection and improving the speed quality the role of data mining is significant. ***Intrusion detection,*** in the business operation of the business they are under the continuous threat of cyber-attacks. The cyber attackers may target that database of an organization, server, web clients and operating system. With the help of data mining the anomalous patterns can be analysed easily and with the careful studying of those patterns it can be figured out whether the concerned organization is under the threat of the cyber attacker or it has been attacked by any of the online entities.

Fraud detection, fraud detection can be challenging these days due to the clever mindset of the fraudulent and cyber criminals. These days the fraudulent and cyber criminals have come with a new and complex pattern for committing the fraud crimes [13]. Data mining techniques can be useful in detection of the intrusion in the same way it will help to detect fraudulent activities. Along with the data mining techniques frauds can also be detected with the help of the technology called machine learning. There is a serious drawback of data mining and machine learning is that it cannot detect the new type of cyber-attacks. The type of cyber-attacks that have been already programmed can easily be detected with the help of machine learning technology. For making and tracking the technology and version of machine learning technique it should be up to date so that it can perform in more and more convenient ways. ***Threat of intelligence gathering,*** the data of cyber security threats are often spread into various areas but the biggest challenge that pose is to find the relevant piece of data for that it can be said that the use of data mining techniques can be significant and helps to cope with the threat of intelligence gathering [14]. The key sector's great challenge can be tactical, operational, and strategic. Tactical threats involve compromise, behavioural changes and failed login attempts.

From studying all these threats that are posed in online information sharing it can be said that the technology of data mining and data analytics can be significant for making the process more and more secure from attacks by the hackers.

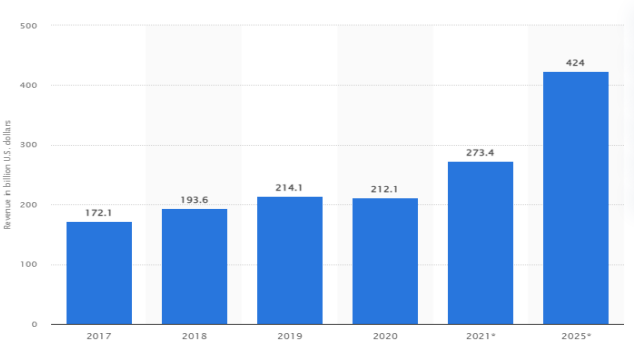

Figure 4: Global enterprises of application software that generates revenue

e-ISSN: 2583-195X

In the government figure above it has been clearly shown that globally the software has risen in the coming year and has been reported at a significant pace. In the year 2017 the total revenue that was generated with the help of application software was around 172.1 billion US dollars [15]. In the next year, 2018 the revenue had risen to 193.6 billion dollars. It cannot be seen as the significant rise but at the same time in the year 2025 which is a prediction the revenue is expected to raise at the mark of 424 billion us dollars and this rise is significant. In the coming years the revenue seems to raise hence the role of data mining also seems to rise. According to statista, in the year 2021 the revenue seems to reach the market to 273.4 billion US dollars. With these statistics it can be assumed that the Covid-19 pandemic of the year 2020 had no impact on revenue generation of the company.

When the pandemic was at its peak a vast majority of the IT professionals seemed to make significant progress in the field technology due to the work from home facilities. A lot of work has shifted to the online mode and most of the tasks started performing via online mode. This change was remarkable but at the same time the threat of cyber-attacks and other threats increased for the company. The use of software applications have also increased in recent years and allow the company where they are a product based company or a service based company to make their application [16]. With the help of software applications many have eased their reach to the customer and now they can sell their products literally to any one at desired price. Moreover, there have been times when it was difficult to find clusters and market their products but now it can be done with the help of application software. There are some of the applications whose performance is a bit slower as compared to the reason can be the bugs and other things.

The revenue will increase in the following years and the rise in the number of users. It can be imagined that in the coming years the scope of development in technology seems to rise. If the technology will be developing in the coming years it is therefore very much possible that the cybercrimes seem to rise as well. Modern techniques are also used by the cyber attacker to hunt useful data of a particular company. Technological development also needs in the field of improving cyber security and for that the data mining is needed to make the process bugs free [17]. The data mining technology has flourished in developed nations like the USA, UK and France and other western nations. There are examples of a lot of data mining companies and those are Oracle data mining, DOMO, rapid miner, Orange data miner and IBM Congnos. These are some of the popular data mining companies; most of them are based in the USA. This popular data mining company basically helps an entity analyse a huge amount of data for home and according to the aims and objectives of the company.

There are some of the issues while spangling the data mining techniques and those issues like the data mining techniques are new and it's costly as well. The techniques of data mining are sought after and therefore there is less expertise that is available in the market. It is due to the need to become an effective data miner that a person should have very good command over any of the programming languages and should have a good command over the data science. For becoming an effective data scientist the knowledge of statistics is needed. The process of data mining is costly but it is effective and the companies are hiring people for this job role [18]. Now coming to the importance of the data manning in the current business it has become significant for a company to sustain in the market and for that it is important to make the process more and more in analysing the customer's interest along with likes and dislikes.

It is a well-known fact that the technology developed in the future will give rise to the rise of information storage. For protecting the online data and improving the online data security it is important to make progress in the process of data mining and secure the data in the online mode.

DISCUSSION

Data mining has gained popularity in recent years after the increase of online data. Online users have significantly increased in the past decade and countries like India and China have the world's largest number of smart phones as well as social media users. Companies like Amazon when it decides to operate its business in India it focuses to make the process as precise as possible and sell the right procedure to the right customer. For this purpose the concept of data mining was used. Amazon basically uses the data of its customers and uses this data to make accurate guesses according to the likes and dislikes of the customers. Moreover, data mining has a lot of benefits and those benefits are like it provides a better understanding of the market situation for a company. Companies can also use those data to predict in a particular country whether their products are laid by the people or not. These are some of the benefits but there are some cons of the process like the lot of information of customers in the market is benefits for the company but if these data are stolen by other entities is the biggest threat. The bio of many off campus personalities is hacked by the hackers and this has been possible because these days there have been a trend of posting everything on social media. Not only this data that are intentionally provided but also the data that are kept confidential are also vulnerable in online mode.

In the result section various graphs have been provided and with the help of those graphs it can be said that the technology is advancing the developed nations. Countries like the USA had been benefited by this technology. As of now it can be said that countries like India and China both have the highest number of smartphone users along with the highest number of social media users. Due to this there is a high possibility that in future the revenue is going to rise further in these countries. According to the first figure it can be figured out that in the future the revenue is going to rise and by the year 2027 the total revenue from the online market will be around 60.09 billion US dollars. Many of the social media as well as those companies that want to take a

e-ISSN: 2583-195X

competitive advantage in the market will take the advantage of such data. The next graph that have been provided in the above section is about the global social network users and according to the graph that have been provided the total no of social users as of the year 2022 is around 4.59 billion and due to such a big rise in the as compared to the previous year it can be predicted that in coming years the social media users is going to hit the mark of 5.85 billion users by the end of the year 2027.

Due to this enormous rise in the number of social media users there is a high probability that data that will be created in the coming years will be enormous. Data security will be a concern for all of us hence by making the process safe for the online user's information security will be needed. For that purpose, using the data mining techniques all the information that has been created online will be taken care of.

CONCLUSION

In this study it discussed in what ways data mining can be useful as well as up to what extent this technique is reliable when it comes to information security. In the introduction part a brief background has been provided regarding the research topic. The aims and objectives have been discussed later in this passage after that a material and methodology section have been provided. In this section the research methodology that has been followed has been discussed by explaining the data collection method and the data analysis. The data have been collected on the basis of a secondary data collection method. After the collection of data, the collected data is further analysed with the data analysis method. Here thematic data analysis has been done and the themes that will be discussed further will be on the basis of the research topic.

In the next section results have been discussed and that too with the help of adequate graphs and pictures that are relevant to the research topic. The graphs that have been provided in the results have given an estimation that globally how much revenue is often generated with the help of online data. Online data is often created on various online platforms like social media and websites. Graphs have also given a rough idea that in the coming years online data is going to rise and the threat the information security is also going to rise further in future. For that purpose, the technology of data mining can be useful and it will help the company to minimize the risk of data theft and improve the data security.

REFERENCES

[1] Dogan, Alican, and Derya Birant. "Machine learning and data mining in manufacturing." Expert Systems with Applications 166 (2021): 114060.

[2] Bock, Frederic E., et al. "A review of the application of machine learning and data mining approaches in continuum materials mechanics." Frontiers in Materials 6 (2019): 110.

[3] Ricciardi, Carlo, et al. "Using gait analysis' parameters to classify Parkinsonism: A data mining approach." Computer methods and programs in biomedicine 180 (2019): 105033.

[4] Snyder, Hannah. "Literature review as a research methodology: An overview and guidelines." Journal of business research 104 (2019): 333-339.

[5] Braun, Virginia, and Victoria Clarke. "Reflecting on reflexive thematic analysis." Qualitative research in sport, exercise and health 11.4 (2019): 589-597.

[6] Liu, Xinyi, Qunying Huang, and Song Gao. "Exploring the uncertainty of activity zone detection using digital footprints with multi-scaled DBSCAN." International Journal of Geographical Information Science 33.6 (2019): 1196-1223.

[7] Taylor.P . Statista. Big data analytics market size worldwide in 2019 and 2025, (2022). https://www.statista.com/statistics/947745/worldwide-total-data-market-revenue/ Accessed 7 January 2023.

[8] Dixon.S . Statista. Number of global social network users 2017-2027, (2022). https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ Accessed 7 January 2023.

[9] United Nations. World population to reach 8 billion on 15 November 2022, (2022). https://www.un.org/en/desa/world-population-reach-8-billion-15-november-2022 Accessed 7 January 2023.

[10] Alexei, Lachi Arina, and Anatol Alexei. "Cyber security threat analysis in higher education institutions as a result of distance learning." International Journal of Scientific and Technology Research 3 (2021): 128-133.

[11] Zhang, Chaobo, et al. "An improved association rule mining-based method for revealing operational problems of building heating, ventilation and air conditioning (HVAC) systems." Applied Energy 253 (2019): 113492.

[12] Wang, Chao, et al. "Research on data mining of permissions mode for Android malware detection." Cluster Computing 22.6 (2019): 13337-13350.

[13] Al-Hashedi, Khaled Gubran, and Pritheega Magalingam. "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019." Computer Science Review 40 (2021): 100402.

[14] Wu, Shaofei, Jun Liu, and Lizhi Liu. "Modeling method of internet public information data mining based on probabilistic topic model." The Journal of Supercomputing 75.9 (2019): 5882-5897.

[15] Vailshery.S.L. Statista. Global enterprise application software revenue 2017-2025, (2022). https://www.statista.com/statistics/247554/global-enterprise-application-software-revenue/ Accessed 7 January 2023.

[16] Rahman, Hamirahanim Abdul, Jinsoo Park, and Jihae Suh. "Use of software agent technology in management information system: a literature review and classification." Asia Pacific Journal of Information Systems 29.1 (2019): 65-82.

[17] 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.

[18] Huber, Steffen, et al. "DMME: Data mining methodology for engineering applications–a holistic extension to the CRISP-DM model." Procedia Cirp 79 (2019): 403-408.