ISSN: 0974-5823 Vol. 7 No. 12 December, 2022

International Journal of Mechanical Engineering

Critical Review on Novel Methods for Preserving Data Mining

Ashoktaru Pal[1], Dr. Lokendra Singh Songare [2]

[1]Research Scholar, Department of Computer Science & Engineering, Dr. A.P.J. Abdul Kalam University, Indore M.P., India.

[2] Research Guide, Department of Computer Science & Engineering, Dr. A.P.J. Abdul Kalam University, Indore, M.P., India.

Abstract:

Digitization has led to an increase in the amount of data being gathered by businesses and people. Big data clusters are used to store massive amounts of data, yet the great majority of organisations have taken no security precautions. Confidentiality of this information is critical. Recently, privacy-preserving data mining has become a hot study subject for the protection of sensitive information on the internet. Data mining with privacy preserved may be accomplished using a wide range of computational approaches. So, in this article critical review on novel methods for preserving data mining has been discussed.

Keywords: Data, Mining, Novel

INTRODUCTION:

Different application domains employ the Knowledge Discovery Process (KDP) or Knowledge Discovery in Databases (KDD) to find new knowledge. Non-trivial extraction of meaningful and hidden information from datasets is another definition. This strategy is used by business analysts, medical scientists, researchers, defense analysts, socialists, and economists to create computer-based strategic judgments. Selection, pre-processing, transformation, mining, and interpretation are all parts of KDP. KDD relies on data mining as its foundation. Algorithms for data exploration, model development, and discovery of previously undiscovered patterns are devised. Use this model to comprehend the data, analyze and forecast the future of the data.

REVIEW OF LITERATURE:

Rathod Shilpa (2020) When it comes to data mining, the ability to protect sensitive data for analysis has made privacy preservation in Data Mining increasingly prominent and popular. Many industries, including healthcare, have generated vast amounts of data in the last decade, and it is critical that this data be analysed and mined for useful information. As an example, a patient's medical records and health test data can be integrated in order to determine the link between an abnormal test result and a condition. By utilising association rule mining on this data, new insights on disease prevention can be gleaned. It is critical that data privacy and security be maintained during the association rule mining method, since the company's sensitive information should not be compromised. In this research, we present an effective method for preserving privacy while mining association rules. As a healthcare datasets-focused work, our method might be applied to a wide range of other fields.

Ritu Ratra (2020) Many institutions, including as hospitals, insurance firms, banks, and the stock market create enormous amounts of data at an ever-increasing rate. Using cutting-edge digitalization technology, this is made possible. Electronic gadgets create enormous amounts of data, as is well-known. Processing of this data might aid in making decisions. In the event of a privacy infringement, data analytics is a risky endeavour. While data analytics is unquestionably beneficial in the decision-making process, it will inevitably raise severe issues of privacy. When it comes to data analytics, protecting people's privacy has become the most critical responsibility. Many challenges to privacy are discussed in this research. Limitations of privacy-preserving techniques and models are also

explored. PPDM algorithms play an increasingly important role today. PPDM strategies for protecting individual privacy have unquestionably risen in popularity in our day and age. Cryptography, secure sum methods, perturbation, and k-anonymity are only a few examples. PPDM research is the primary emphasis of this article. This study will help readers better grasp the many difficulties that PPDM faces. It will also be beneficial to understand and implement the most appropriate strategy for diverse data sets. –

Ritu Ratra and Preeti Gulia (2020) Many institutions, including as hospitals, insurance firms, banks, and the stock market create enormous amounts of data at an ever-increasing rate. Using cutting-edge digitalization technology, this is made possible. Electronic gadgets create enormous amounts of data, as is well-known. Processing of this data might aid in making decisions. There is a risk, though, that data analytics might violate privacy laws. While data analytics is unquestionably beneficial in the decision-making process, it will inevitably raise severe issues of privacy. When it comes to data analytics, protecting individual privacy became the most critical and required duty. Many challenges to privacy are discussed in this research. Limitations of privacy-preserving techniques and models are also explored. PPDM algorithms play an increasingly important role today. In today's world, there are no doubt a lot of PPDM approaches in use to protect individual privacy. Cryptography, secure sum methods, perturbation, and k-anonymity are only a few examples. PPDM research is the primary emphasis of this article. This study will help readers better grasp the many difficulties that PPDM faces. It will also be beneficial to understand and implement the most appropriate strategy for diverse data sets. –

Stephen, Oloo Ajwang (2020) Mobile platforms and linked devices have transformed data collection, storage and mining in agriculture because of their capacity to collect large volumes of data. To make decision-making easier and more efficient, these new technologies are being used in a variety of fields, from agriculture to the home, to the workplace, and everywhere in between. Nevertheless, on board data mining has been hampered owing to the inherent limitations of mobile platforms including sluggish processors and tiny displays to show the results due to low-bandwidth networks. In addition, mobile devices run on a variety of platforms, making it difficult to integrate server applications. A service-oriented architecture (SOA) based on web services, as well as an artificial neural network (ANN), were proposed in this research to address these issues, and they were used to enable the mobile data mining of huge agronomic data sets and the prediction of yield and weather trends. After a thorough examination of the current mobile data mining architecture, the design was presented. SOA was the perfect choice since it leverages web services to promote interoperability between clients and server applications regardless of the platforms they are executed on, thereby giving mobile devices with data mining capabilities. Based on the SO-M-Miner model, this study presents a 7-layer architectural design. SMS gateway, data client, network, web service, database and ODBC connector were some of the components of the system.

Chamikara et. al., (2019) PABIDOT is an effective and scalable non-reversible perturbation approach for preserving privacy in huge data through optimal geometric transformations. Five different classifiers were used to test the perturbation 37 algorithm's capacity to withstand a variety of attacks, as well as its accuracy and efficiency. According to the results, this technique outperforms prior privacy-preserving algorithms in terms of scalability, accuracy, and speed to perform large-scale privacy-preserving categorization of datasets.

Yao et. al., (2019) To protect the privacy of personal data by interrupting the release of published health care data (PPHR). The initial step in the suggested effort was to find the fundamental sequences that might be used to quickly identify those unique texts. That's when the authors allow for tampering with these sequences by adding or removing those points, as long as the L-diversity is consistent with published data. The effectiveness of the suggested anonymization technique is evaluated using a comprehensive set of real-life healthcare datasets. When all was said and done, the new system was found to be more useful and more private.

Zhang et. al., (2019) that the combination of a balanced iterative reduction and clustering utilising hierarchy (BIRCH) and a Maximum Delay Anonymous Clustering Function tree data publication system might be used to preserve the privacy of private data shared on social networking sites. On-and off-line components make up the integrated technique that has been offered. The Anonymous Clustering Feature

(ACF) tree and the feature vector were used to quickly cluster data streams in the Online process and classify the data collection into discrete beginning groups. K-anonymous The diversity system is used to anonymize the groups when the group's leaf node size exceeds the given threshold. Offline processing is required for data anonymization. Measurements of the distance and lack of interest between data are used to remove anomalous numbers.

Varsha Patel (2019) Data mining for a shared data storage system that preserves privacy is known as privacy-preserving data mining. To maintain data sensitivity and privacy, this method uses pooled data to provide an accurate analysis. The primary focus of this study is on privacy-preserving data mining, as well as a new system implementation. A variety of newly developed PPDM-based data mining algorithms are used in this area. A novel data mining approach is also proposed in this study to help construct an accurate and effective PPDM model, as well. Finally, the study outlines how the proposed research may be expanded in the future.

Anshu (2019) Data mining techniques and applications were explored in the work titled "Review Paper on Data Mining Techniques and Applications." Data mining is a technique for discovering patterns and important information buried in large datasets. Data mining is a novel approach that enables businesses to anticipate future trends and behaviours and to make data-driven decisions. This study's goal is to illustrate how data mining may help decision-makers make better choices. Any business with a huge amount of data may greatly benefit from data mining. When data mining is applied on regular databases, the performance improves. Profits rise as a result of better judgments made with data mining's assistance. An in-depth look at the various data mining techniques and how they may be used to various organisations is provided in this article.

Vahedifard, Farshid et. al., (2019) DROVE is one of the most complete physical testing databases for off-road vehicles in existence. This document provides an update from a multi-year research initiative to establish and improve DROVE. Data from published and unpublished reports for laboratory and field experiments spanning decades of wheeled and tracked vehicle testing on various soil types were mined to construct and extend DROVE over a multiyear period. Over 8,000 entries from existing archives of laboratory and field tests of wheels running in loose sand and high plasticity clay were included in the database's initial version (DROVE 1.0). Results from DROVE 1.0 contained data from a wide range of wheel sizes and inflation pressure tests under a variety of loading scenarios. It is anticipated that the introduction of DROVE 3.0 will give terramechanics researchers with hundreds of digital files containing mobility data, as compared to the 294 test results provided by DROVE 2.0 on fine-grained soils. Drawbar pull, torque, traction, motion resistance, sinkage, and wheel slip are among the traction performance characteristics included in the DROVE framework. For the mobility community, the DROVE platform serves as a resource for evaluating and improving existing soil mobility algorithms, as well as for developing new algorithms to evaluate wheel performance. The authors compared DROVE's algorithms to those already described for wheeled vehicle mobility modelling in the Vehicle Terrain Interface (VTI). For various soil types, a comparison of anticipated and measured performance metrics is shown. An analysis of published data, its limits, and the data's compatibility with force prediction methods such as VTI is provided in this article.

Martyniuk, Hanna et. al., (2019) Information about social media and data mining is included in this study. Facebook has an 85 percent share of the world's internet users, making it the world's most popular social networking site. Even if you don't use a search engine, using data mining techniques on large social media data sets can help improve search results, help businesses target specific customers, help psychologists understand behaviour, help sociologists better understand society's structure, and even help all of us avoid spam. There are examples of the most frequent data mining applications connected to social networking sites. Other data mining approaches and their respective lists have also been provided by the authors. Social network data must be protected against unauthorised access. Recent studies have demonstrated that even when data has been anonymized, modern data analysis tools can still disclose personal information, underscoring the need of privacy protection. There is a wide range of social network dangers depicted. For example, they teach how to protect yourself against identity theft and cyberbullying on social networks by backing up your data.

Upadhyay et al. (2018) Using data partitioning and three-dimensional rotations, the author has proposed a method of geometric data perturbation

(GDP). As a result of this method, the characteristics are broken down into groups and then rotated around the axis in different directions. Data mining methods such as classification and clustering are invariant to geometric perturbation, which makes them ideal for machine learning.

Fan et al. (2018) Using unsupervised data analytics, he has revealed the results of a thorough investigation of the current state of large-scale data-driven building mining. Unsupervised analytics, which are widely used, are described in terms of representations and applications of knowledge.. In this multi-disciplinary discipline, challenges and prospects are enhanced as a guide for prospect research.

Kornilakis et al. (2018) Using an incognito browser for social media was recommended OSN pages may now be accessed without the danger of revealing sensitive information about users, such as their search habits or preferences for a social media account. There will be no monitoring methods in place to restrict how users search for material that cannot be made publicly available on the server while they access sensitive information in a private mode of functioning (incognito). Users using incognito browsers will enjoy minimum latency and low bandwidth needs.

Amiri Fatemeh (2018) It is the goal of this study to find and offer relevant ways to overcome the privacy gaps in certain data mining activities with three key objectives. As a whole, the goal is to maintain the processes running as efficiently as possible while attempting to increase privacy. With regard to protecting user privacy in big data, we show that a machine learning approach is a viable option. Data mining can benefit greatly from our approach, which includes numerous typical efficient methods for privacy-preserving calculations. For e-business application clustering in remote contexts, a set of better algorithms based on soft computing that preserve privacy is the intended consequence and contribution of this article. By demonstrating how soft computing approaches may lead to new outcomes for privacy protection as well as maintaining performance and accuracy throughout routine operations, the suggested model shows that soft computing methods are very useful in online commerce.

Aghasian et al. (2018) have recommended ways to keep social media platforms like Facebook, Instagram, and Twitter from misusing user data. Users must protect their personal information from prying eyes because of a rise in privacy concerns. They've come up with a novel privacy-protection mechanism in an effort to keep confidential information from being leaked. The author outlined the kind of knowledge that should be disseminated to others. The Bernstein polynomial theorem is used in the suggested technique to help consumers better comprehend the types of data that might have an impact on their security. After then, a new model is used to anonymize the data of people who are connected to someone on social media.

Dr. Chahal (2018) The term "data mining" refers to the act of using an existing database to discover new information, rules, and patterns. Unstructured, massive amounts of data are mined for their knowledge and then analysed. Knowledge discovery is another name for this process. The resulting data and trends are an invaluable resource for strategic decision-making and company planning. Users' private information should not be exposed by the results of Data Mining. Privacy Preserving Data Mining Strategies are the techniques used to protect data from malicious users. Data mining architecture that protects privacy is examined in this research. The bands of Privacy Preserving Objectives are also shown. The study summarises various privacy-preserving strategies. Researchers and scientists in the field of privacy-preserving data mining are the intended audience for this study.

Kaya Keleş, Mümine. (2018) Increasingly complex, non-significant, confusing, massive, and raw data has made it increasingly difficult to obtain trustworthy information in recent years. As the amount of data grows, so does the demand for precise and dependable data analysis. It is possible to obtain trustworthy and useful information using the Data Mining Method, which is a statistical application. Information on how data mining is utilised in various industries was presented in this study, as well as a review of the literature in the fields of banking and finance and information technology as well as health care and public administration. We want to fill a vacuum in the literature and bring innovation to these areas by doing this research.

Froemelt, Andreas et. al., (2018) Household consumption is a major driver of the economy and may be viewed as ultimately accountable for environmental consequences that occur along the life cycle of products and services. It is necessary to study home behaviour patterns and environmental implications in order to devise

targeted environmental solutions since purchasing decisions are made at the household level and are heavily influenced by behaviour. We suggest a new technique to capturing the variety of lifestyle-induced environmental consequences in order to promote successful environmental policymaking. A Ward-clustering is used on a pre-conditioning self-organizing map to generate a two-tiered clustering of lifestyle archetypes. A hybrid life cycle assessment methodology is then used to evaluate the environmental implications of certain archetypal behaviour. A worldwide picture of consumption may be obtained in the Swiss Household Budget Survey using this method, but it also shows that distinct archetypes can be found within the same socioeconomic groups. It is possible that the presented technique might be beneficial for a better understanding of consumption patterns and the development of environmentally-targeted policies for specific consumer groups in the future, given the advent of archetypes that deviate from overall macro-trends.

Arumugam, G., & Sulekha, V. J. V (2017) Every record identified as; preference enforced individual ranking based micro aggregation with optimal noise addition technique was presented as a strategy. (P-IRON). Prior to implementing data protection techniques, the absence of value and knowledge is prioritised. Despite the fact that the evidence was compromised, the results demonstrated that the proposed technique for preventing the transmission of secret material worked.

Senosi & Sibiya, A. (2017) In their review of Privacy Preservation Data Mining, the researchers have proven (PPDM). The classification of PPDM techniques and the description of techniques were also detailed in this study. Using data from the literature, the author created a taxonomy of these strategies and tracked their performance on several assessment metrics. However, this work does not necessitate any more study, given the present standards used by academics are not standardised.

Sriramoju (2017) The author discussed the privacy preservation model's impact on big data in light of the author's point of view. The author also discusses the benefits and drawbacks of various frameworks for maintaining the confidentiality of data. A correlation was found between execution time, implementation difficulty, and the practical value of the resulting data.

Pensa, R.G., and Di Blasi, G. (2017) An assessment technique developed for strengthening privacy in a single source, such as social networking sites, has been detailed. Two main components of this research have been demonstrated in the form of a data protection ranking that could be monitored to warn every person disclosed for privacy breaches, and an inactive learning method that can be helpful for users to maximise the security through minimising the amount of manual operations that are necessary. The authors of this study focused on images shared on social media.

Ma et al. (2017) The de-anonymization of social networks may be done using a random technique. As a result of social networking sites sharing user information with investigators, advertisers, and project managers, user privacy is threatened. By removing personal information, such as a user's identity card and email address, this data has been made anonymous. However, the user's personal details could not yet be sufficiently secured. With spectral partitioning, enormous sparse social networks were divided into several little sub-graphs. The anonymous network and auxiliary network utilising the suggested technique provides random forest categorization for candidate node pairs that are matched. The results showed that efficiency and accuracy had improved.

CONCLUSION:

Data mining is the process of extracting information from a large amount of data. In this phase, the data that will be utilized in the KDP is determined. It entails doing research to see what resources are accessible and gathering any extra information that may be required. It combines all of the information needed to uncover new information into a single database. In addition, it contains the characteristics that will be taken into account during the procedure. To build models, we need this evidence. If certain essential characteristics are absent, the process as a whole may not work appropriately. At this point, you have to decide on the strategy you'll use to look for patterns. For example, when comparing accuracy and understandability, neural networks come out on top, whereas decision trees come out on top when it comes to the latter. Meta-learning strategies can be implemented in a variety of different ways. A Data Mining algorithm's success or failure in a given problem is the focus of meta-learning. There are parameters and learning strategies

for every algorithm. As a result, this technique aims to identify the best Data Mining algorithm for a given situation.

REFERENCES:

Aghasian, E., Garg, S. and Montgomery, J., 2018. A privacy-enhanced friending approach for users on multiple online social networks. Computers, 7(3), p.42.

Anshu, Review Paper on Data Mining Techniques and Applications (MARCH 31, 2019). International Journal of Innovative Research in Computer Science & Technology (IJIRCST), Volume-7, Issue-2, March 2019, Available at SSRN: https://ssrn.com/abstract=3529347.

Arumugam, G., &Sulekha, V. J. V. (2017, August). P-IRON for Privacy Preservation in Data Mining. In International Conference on Knowledge Management in Organizations, Springer, Cham, pp. 410-423.

Chahal, Dr. (2018). Privacy Preserving in Data Mining. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 3. 2079-2083.

Chamikara, M. A. P., Bertok, P., Liu, D., Camtepe, S., & Khalil, I. (2019). Efficient privacy preservation of big data for accurate data mining. Information Sciences.

Fan C, Xiao F, Li Z, Wang J. Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. Energy and Buildings. 2018 Jan 15;159: pp. 296-308.

Fatemeh Amiri, A Machine Learning Approach for Privacy-preservation in Ebusiness Applications, In Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018) - Volume 2: SECRYPT, pages 443-452.

Froemelt, Andreas & Dürrenmatt, David & Hellweg, Stefanie. (2018). Using Data Mining To Assess Environmental Impacts of Household Consumption Behaviors. Environmental Science & Technology. 52. 10.1021/acs.est.8b01452.

Kaya Keleş, Mümine. (2018). An overview: the impact of data mining applications on various sectors. Tehnički Glasnik. 11.

Kornilakis, A., Papadopoulos, P. and Markatos, E., 2018, September. Incognitus: Privacy-Preserving User Interests in Online Social Networks. In International Workshop on Information and Operational Technology Security Systems, Springer, Cham pp. 81- 95.

Ma, J., Qiao, Y., Hu, G., Huang, Y., Sangaiah, A.K., Zhang, C., Wang, Y. and Zhang, R., 2017. De-anonymizing social networks with random forest classifier. IEEE Access, 6, pp.10139-10150.

Martyniuk, Hanna & Lazarenko, Serhii & Kozlovskiy, Valeriy & Balanyuk, Yuriy & Yakoviv, Ivan & Skladannyi, Pavlo. (2019). Data Mining Usage for Social Networks.

Oloo Ajwang, Stephen. (2020). Service-oriented Data Mining Architecture for Climate-Smart Agriculture. American Journal of Data Mining and Knowledge Discovery. 5. 10.11648/j.ajdmkd.20200501.11.

Pensa, R.G. and Di Blasi, G., 2017. A privacy self-assessment framework for online social networks. Expert Systems with Applications, 86, pp.18-31.

Ratra, Ritu & Gulia, Preeti. (2020). Privacy Preserving Data Mining: Techniques and Algorithms. International Journal of Engineering Trends and Technology. 68. 56-62. 10.14445/22315381/IJETT-V68I11P207.

Ritu Ratra, Preeti Gulia "Privacy Preserving Data Mining: Techniques and Algorithms" International Journal of Engineering Trends and Technology 68.11(2020):56-62.

Senosi, A., &Sibiya, G. (2017, September). Classification and evaluation of privacy preserving data mining: a review. In 2017 IEEE AFRICON, IEEE, pp. 849-855.

Shilpa Rathod, Survey on Privacy Preserving Data Mining Techniques, Volume 09, Issue 06, June 2020.

Sriramoju, S. B. (2017). Analysis and Comparison of Anonymous Techniques for Privacy Preserving in Big Data. International Journal of Advanced Research in Computer and Communication Engineering, 6(12), 2278-1021.

Upadhyay, S., Sharma, C., Sharma, P., Bharadwaj, P., & Seeja, K. R. (2018). Privacy preserving data mining with 3-D rotation transformation. Journal of King Saud University-Computer and Information Sciences, 30(4), 524-530.

Vahedifard, Farshid & Howard, Isaac & Borazjani, Arman & Mason, George & Priddy, Jody. (2019). Update From A Multi-Year Data Mining Effort On Drove: Database Records For Off-Road Vehicle Environments.

Varsha Patel, A Study of Privacy Preserving Data Mining and Techniques, International Research Journal of Engineering and Technology (IRJET). Volume: 06 Issue: 03 Mar 2019.

Yao, L., Wang, X., Chen, Z. and Wu, G., 2019, April. Privacy Preservation in Publishing Electronic Health Records Based on Perturbation. In International Conference on Security and Privacy in New Computing Environments Springer, Cham pp. 125-140.

Zhang, Z. and Gupta, B.B., 2019. Social media security and trustworthiness: overview and new direction. Future Generation Computer Systems, 86, pp.914-925.