JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 701 - 708
https://publishoa.com
ISSN: 1309-3452

A Survey on Privacy Preserving Data Mining its Related Applications in Health Care Domain

S.Joseph Gabriel[1]*, Dr.P.Sengottuvelan[2]*

[1*]Research Scholar,Department of Computer Science, Periyar University, Salem
talk2anette67@gmail.com
[2]Associate Professor, Department of Computer Science, P.G Extension Centre,
Periyar University, Dharmapuri.
sengottuvelan@gmail.com
(Corresponding Author:S.Joseph Gabriel)

ABSTRACT

Data Mining is a technical process which involves the conversion of a raw data into a useful data. There is generally certain sequence of steps involved in obtaining a useful information. One of the important usages of the data mining technique is to identify certain useful patterns that can be utilized to solve certain real-world problems. Utility mining is a domain of data mining which is mainly used to identify the high utility based item sets which are considered based on certain parameters like quantity, cost, profit, affordability and certain preferences of the users. Utility mining impacts various domains like online commerce industry, medical domains, biomedical applications, finance, marketing and its allied domains. In this survey we discuss about the general structure in the pattern mining and various algorithms being applied to the privacy preservation like the randomization, k-anonymity and certain other hybrid techniques that is applied to the health care domain for preserving the privacy even though when the data is stored in the cloud and we consider various real-world problems and discuss about the open challenges in this research domain.

Keywords- Data Mining, privacy preserving data mining, K-anonymity, randomization, cloud, health care domain, Utility based item mining.

INTRODUCTION

Data mining is increasingly gaining its popularity in the industrial domains. But the applications of the correct data mining technique to the problem are the key ingredient to the success. The important theme of data mining is to extract information at a high-level from the data which is raw and available in surplus. It primarily targets a data analysis to obtain information and it can be used for large and even small volume of data and can be useful for taking decisions that are strategic. Data mining can be thought of as an algorithm which consumes data as the input and produces patterns like item sets, rules for classification, and rules for Association. But due to data mining there is a threat to the privacy of the data because of the collection of data and how securely the data is stored and retrieved without losing privacy. Because most of the organizations are depended on the cloud based services for the storage of data the task of preserving the privacy is very crucial. Most of the methods that are used for preserving privacy use some type of data transformations. We discuss the various algorithms for high utility item set mining and privacy preserving algorithms and we discuss about certain open research problems in this area.

TECHNIQUES IN DATA MINING

The Data Mining Techniques are broadly classified into various process like Classification, clustering, Regression, Association Rules, Outlier detection, Sequential Patterns, prediction process.

Clustering

Clustering is a process of separation of the objects that are connected in groups. Clusters are used to model the data. Clustering is mainly helpful in the domains of data mining like text mining, retrieval of information, web mining, and diagnostics of medical conditions. The analysis of clustering is data mining approach for identifying the data which is similar. Clustering is almost closer in mechanism of a classification where it combines the parts of data into groups based on their similarities.

Classification

Classification with respect to the data mining can be done based on the data source type that is mined where the

classification will be done based on the type of data that is handled, and it can be based on the database that is involved where it can be an transaction based database, classification can also be done based on the type of knowledge to be discovered where steps like discrimination, clustering are involved and it can also be performed based on the other techniques used like neural networks, visualization, genetic algorithms, database based and statistics.

Regression
Regression analysis is used to identify the relationship the variable and other factor. A specific variable property will be defined by the regression. It is a type of modelling and plan. We can use the regression analysis to predict or forecast how the demand changes whether it grows or diminishes according to the needs of the customer.

Association rules
Association rules describe the relationship between two or more items and it is used in the dataset to find some of the hidden details. Association rule mining has a good application needs in the medical datasets.it has three important measurements namely confidence, support and lift.

Patterns in sequential order
It is mainly used to identify the sequential data to identify the sequential pattern. It is mainly used to find some sequences of interest to the user. It is mainly used to identify the similar sequence of transactional patterns.

Outlier detection
Outliers are the unwanted data which are not of much use to the end user and it is very important to remove those outliers for effective prediction of useful patterns. This technique is very much useful for various domains like detection of intrusion, detection of fraud. It is known to be mining of the outlier or also as analysis of the outlier.

DATA MINING IN HEALTH CARE

Different divisions adequately use information mining. It empowers the retail segments to show client reaction and encourages the financial division to foresee client gainfulness. It serves numerous comparative segments, for example, producing, telecom, medical care, car industry, training, and some more.
Mining of information holds extraordinary potential for medical care benefits because of the exponential development in the quantity of electronic wellbeing records. Beforehand Doctors and doctors hold persistent data in the paper where the information was very hard to hold. Digitalization and advancement of new procedures decrease human endeavours and make information effectively assessable. For instance, the PC keeps a voluminous measure of patient information with precision, and it improves the nature of the entire information the executives framework. In any case, the significant test is what medical care administrations suppliers should do to channel all the information proficiently. This is where information mining has demonstrated to be incredibly helpful.
Researchers are using various methodologies like groups, order, choice trees, neural organizations, and time arrangement to distribute research. In any case, Healthcare has reliably been delayed to join the most recent investigation into regular practice.

THE INVESTIGATION FRAMEWORK:

The investigation framework joins the innovation and mastery to collect data, grasp it, and normalize estimations. Conglomerating clinical, persistent fulfilment, money related, and other information into a venture information distribution centre is the establishment of the framework.

The substance framework:
The substance framework incorporates normalizing information work. It applies proof based prescribed procedures to mind conveyance. Researchers make critical disclosures every year about clinical best practice, yet it referenced already, it takes a long effort for these revelations to be fused into clinical practice. A solid substance framework empowers associations to incorporate the most recent clinical adaptation rapidly.

The arrangement framework:
The arrangement framework includes driving change the board over new progressive structures. Especially, it incorporates actualizing bunch structures that engage reliably, undertaking wide arrangement of best practices. It requires a genuine progressive change to drive the appropriation of best practices all through an association.

Use of Data Mining in Healthcare:

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 701 - 708
https://publishoa.com
ISSN: 1309-3452

Information mining has been utilized seriously and broadly by various enterprises. In medical services, information mining is turning out to be more mainstream these days. Information mining applications can staggeringly profit all gatherings who are associated with the medical services industry. For instance, information mining can help the medical care industry in extortion identification and misuse, client relationship the executives, viable patient consideration, and best practices, reasonable medical services administrations. The a lot of information produced by medical services exchanges are excessively unpredictable and colossal to be prepared and examined by ordinary strategies.
Information mining gives the system and methods to change these information into valuable data for information driven choice purposes.

Treatment viability:
Information Mining applications can be utilized to evaluate the adequacy of clinical medicines. Information mining can pass on investigation of which game-plan exhibits powerful by looking at and separating causes, side effects, and courses of medicines.

Medical care the board:
Information mining applications can be utilized to distinguish and follow interminable sickness states and impetus care unit patients, it declines the quantity of clinic confirmations, and supports medical care the executives. Information mining used to dissect monstrous informational collections and measurements to look for designs that may exhibit an attack.

Client relationship the board:
Client and the executives cooperations are essential for any association to accomplish business objectives. Client relationship the board is the essential way to deal with overseeing cooperations between business associations regularly retail parts and banks, with their clients. So also, it is significant in the medical services setting. Client associations may occur through call communities, charging offices.

Misrepresentation and misuse:
Information mining misrepresentation and misuse applications can zero in on unseemly or wrong remedies and extortion protection and clinical cases.

SOME TECHNIQUES OF PRIVACY PRESERVATION

Randomization method

Randomization method is one of the important approaches for preserving privacy in data mining. In the randomization process a random noise will be added to the data in such a way that the normal behaviour of the data will be changed but the normal behaviour of the data will be obtained when the noise is once again removed.
Applications of randomization

Randomization approach can be utilized in various data mining approaches. It is mainly suitable for classification type problems and can be utilized with the classification problems which also provides privacy preservation. Randomization techniques can be used with other technologies like online Analytical processing server (OLAP). Randomization method can be used with other approaches like the mixture models to provide the same task of preserving the privacy. Here we discuss certain research works carried out by using the randomization approach

Huang, Z., & Du, W. (2008, April) in their work uses randomized response technique for preserving the privacy of the categorical data they proposed a method to quantitatively measure the privacy and utility. They used the evolutionary multi-objective technique to find the matrix for the randomized response approach.

Zhu, Y., & Liu, L. (2004, August) in their approach developed a scheme for randomization for privacy preserving in the estimation of density and they proposed a framework for randomization using the mixture models. The effect of randomization on data mining is measured by the amount of information which is lost and the degradation of the performance and they measure by simulations and demonstrated how the privacy is preserved.

Erlingsson, Ú., Pihur, V., &Korolova, A in their work developed a technology called Randomized aggregatable privacy preserving ordinal response is used for obtaining the statistics of crowdsourcing from client software. This methodology allows a large amount of client data to be studied but without allowing to look into the details of the individual trees. It provides high utility analysis of the data which is collected by the user. This method provides a good amount of privacy preservation and it is applied to real as well as synthetic data for testing its credibility.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 701 - 708
https://publishoa.com
ISSN: 1309-3452

Lohiya, S., &Ragha, L in their work randomized the original data and then the generalization approach is applied over the data which is randomized. They found that their technique protects the data and preserves the privacy with a good accuracy and there is no much loss in the information and makes the data usable.

Lin, J. L., & Liu, J. Y. C in their research work understands about the problem of privacy-preserving Association rule mining. They designed a fake randomization method for the transaction for protecting the privacy of the data. They proposed a algorithm for once again obtaining the itemset which occurs frequently from both real transactions and fake transactions. This work also provides some improvement in the mined results.

Privacy Preserving Data mining

Privacy preserving is an important concern for telling whether the data mining is successful or not. Nowadays every individual is aware about the privacy and they are little reluctant to share their important and sensitive data because of the fear of privacy breach and even if they are mandatorily required to share certain important data it is very much essential to preserve the privacy of that data.

The proposed privacy preserving data mining algorithm is considered to be useful when it performs well in-terms of data usage, uncertainty levels to data mining algorithms, performance. In the data mining process the data is obtained from various organization and its format is changed for further analysis purposes and after that the data is stored in the databases and the data mining algorithms will be applied on it for getting the knowledge.

PPDM will in general change the first information with the goal that the aftereffect of information mining assignment ought not to resist security requirements. Following is the rundown of five measurements based on which distinctive PPDM Techniques can be characterized (I) Data appropriation (ii) Data adjustment (iii) Data mining calculations (iv) Data or rule covering up (v) Privacy protection

The main measurement is identified with dispersion of information: Centralized or Distributed. Dispersed information can be evenly or vertically appropriated. Flat circulation alludes to situations where various records dwell in better places while vertical conveyance alludes to situations where all estimations of various credits live in better places. The subsequent measurement alludes to the alteration of unique estimations of information that are to be delivered for information mining task. Change is done utilizing bother, obstructing, total, blending, trading or examining or any mix of these. The third measurement is that of information mining calculations. The information mining calculations are applied on the changed information to get valuable pieces of data that were covered up already. The fourth measurement alludes to whether the crude information or amassed information ought to be covered up. The fifth and the last measurement alludes to the strategies that are utilized for securing protection. In light of these measurements, diverse PPDM methods might be grouped into following five classifications

1. Anonymization based PPDM 2. Annoyance based PPDM 3.Randomized Response based PPDM 4.Condensation methodology based PPDM 5.Cryptography based PPDM we examine these in detail in the accompanying subsections.

ANONYMIZATION BASED PPDM

The essential type of the information in a table comprises of following four kinds of traits.
(I) Explicit Identifiers is a lot of qualities containing data that distinguishes a record proprietor expressly, for example, name, SS number and so forth
(ii) Quasi Identifiers is a lot of characteristics that might recognize a record proprietor when joined with freely accessible information.
(iii) Sensitive Attributes is a lot of traits that contains touchy individual explicit data, for example, sickness, compensation and so on
(iv) Non-Sensitive Attributes is a lot of qualities that makes no issue whenever uncovered even to dishonest gatherings.

Anonymization alludes to a methodology where personality or/and delicate information about record proprietors are to be covered up. It even expect that sensitive information ought to be held for examination. Clearly express identifiers ought to be taken out yet there is a peril of protection interruption when semi identifiers are connected to openly accessible information. Such assaults are called as connecting assaults. For instance credits, for example, DoB, Gender, Race, SSN number, Zip are accessible out in the open records, for example, citizen list. Such records are accessible in clinical records additionally, when connected, can be utilized to gather the character of the relating individual with high likelihood. The semi identifiers like DoB, Gender, SSN, Race, Zip and so forth are accessible in clinical records and furthermore in citizen list that is freely accessible. The unequivocal identifiers like Name, SS number and so forth have been eliminated from the clinical records. All things considered, character of individual can be

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 701 - 708
https://publishoa.com
ISSN: 1309-3452

anticipated with higher likelihood. Sweeney in their work proposed k-obscurity model utilizing speculation and concealment to accomplish k-secrecy for example any individual is discernable from in any event k-1 different ones regarding semi identifier trait in the anonymized dataset. As it were, we can characterize a table as k-unknown if the Q1 estimations of each tuple are indistinguishable from those of in any event k1 different tuples. Supplanting an incentive with less explicit however semantically consistent esteem is called as speculation and concealment includes impeding the qualities. Delivering such information for mining lessens the danger of recognizable proof when joined with publically accessible information. Yet, simultaneously, precision of the applications on the changed information is decreased. Various calculations have been proposed to actualize k-obscurity utilizing speculation and concealment lately. Constraints of the k-obscurity model originate from the two suppositions. To begin with, it might be hard for the proprietor of an information base to figure out which of the traits are accessible or which are not accessible in outer tables. The subsequent impediment is that the k-namelessness model expect a specific technique for assault, while in genuine situations; there is no motivation behind why the aggressor ought not attempt different strategies. Nonetheless, as an examination bearing, k-obscurity in blend with other security saving strategies should be explored for distinguishing and in any event, obstructing k-anonymity infringement.

ANNOYANCE BASED PPDM

Perturbation has a long back history, being utilized in factual divulgence control as it has an inborn property of effortlessness, proficiency and capacity to safeguard measurable data.. The bothered information records don't relate to certifiable record proprietors, so the assailant can't play out the touchy linkages or recoup delicate data from the distributed information. In irritation approach, a record delivered is manufactured for example it doesn't compare to genuine elements spoke to by the first information. Along these lines the individual records in the irritated information are pointless to the human beneficiary as just measurable properties of the records are protected. Bother should be possible by utilizing added substance or information trading or engineered information age. Since strategy doesn't remake the first qualities yet just the appropriations, new calculations are to be created for mining of the information. This implies another appropriation-based mining calculations should be produced for every individual information 28 issues like characterization, bunching or affiliation rule mining. For instance, Agrawal built up another circulation based information digging calculation for the grouping issue. While a few methodologies have been produced for appropriation based digging of information for issues, for example, affiliation rules and grouping, obviously utilizing dispersions rather than unique records confines the scope of algorithmic procedures that can be utilized on the information. Kantarcioglu and Clifton and Rizvi and Harista have created strategies to save protection of affiliation rule mining. In the irritation approach, any circulation based information mining calculation works under a certain supposition to treat each measurement freely. Pertinent data for information mining calculations, for example, grouping stays covered up in between property relationships. For instance, the order strategy utilizes dissemination based simple of single-attribute split calculation. Nonetheless, different methods, for example, multivariate choice tree calculations can't be adjusted appropriately to work with the annoyance approach. This is on the grounds that the bother approach treats various credits autonomously. Consequently the dispersion based information mining calculations have an intrinsic drawback of loss of certain data accessible in multidimensional records. Another part of security saving information mining that deals with the detriments of bother approach is cryptographic procedures that we talk about in one of the accompanying areas.

RANDOMIZED RESPONSE BASED PPDM

Basically, randomized reaction is factual procedure acquainted by Warner with take care of a review issue. In Randomized reaction, the information is mixed so that the focal spot can't tell with probabilities better than a pre-characterized limit, regardless of whether the information from a client contains honest data or bogus data. The data got from every individual client is mixed and if the quantity of clients is essentially huge, the total data of these clients can be assessed with acceptable measure of precision. This is helpful for choice tree order since choice tree arrangement depends on total estimations of a dataset, as opposed to singular information things. The information assortment measure in randomization technique is done utilizing two stages. During initial step, the information suppliers randomize their information and send the randomized information to the information recipient. In second step, the information collector remakes the first dispersion of the information by utilizing a dissemination reproduction calculation. Subsequently, the randomization strategy can be executed at information assortment time. It doesn't need a believed worker to contain all the first records so as to play out the anonymization cycle. The shortcoming of a randomization reaction based PPDM method is that it treats all the records equivalent independent of their nearby thickness. This prompts an issue where the exception records become more powerless to ill-disposed assaults when contrasted with records in more thick areas in the information. One answer for this is to be unnecessarily more forceful in adding clamor to all the records in the information. Be that as it may, it lessens the utility of the information for mining purposes as the reproduced circulation may not yield brings about similarity of the motivation behind information mining. The randomized reaction approach has stretched out its qualities to various information mining issues. The randomization approach has additionally been reached out to different applications, for example, OLAP, and SVD based communitarian separating.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 701 - 708
https://publishoa.com
ISSN: 1309-3452

Buildup approach based PPDM

Condensation approach develops compelled groups in dataset and afterward creates pseudo information from the insights of these bunches. It is called as build up as a result of its methodology of utilizing consolidated insights of the groups to produce pseudo information. It develops gatherings of non-homogeneous size from the information, with the end goal that it is ensured that each record lies in a gathering whose size is at any rate equivalent to its namelessness level. Accordingly, pseudo data is produced from each gathering in order to make an engineered informational index with a similar total dispersion as the first information. This methodology can be adequately utilized for the issue of grouping. The utilization of pseudo-information gives an extra layer of assurance, as it gets hard to perform ill-disposed assaults on manufactured information. Besides, the total conduct of the information is saved, making it valuable for an assortment of information mining issues. This methodology helps in better protection conservation when contrasted with different procedures as it utilizes pseudo information instead of changed information. In addition, it works even without updating information mining calculations since the pseudo information has a similar organization as that of the first information. It is exceptionally Original Dataset Randomized Dataset Original Distribution 29 powerful if there should arise an occurrence of information stream issues where the information is profoundly unique. Simultaneously, information mining results get influenced as enormous measure of data is lost due to the buildup of a bigger number of records into a solitary factual gathering substance.

Cryptography based PPDM

Consider a situation where numerous clinical establishments wish to lead a joint exploration for some shared advantages without uncovering pointless data. In this situation, research with respect to manifestations, finding and medicine dependent on different boundaries is to be directed and simultaneously security of the people is to be ensured. Such situations are alluded to as dispersed figuring situations. The gatherings engaged with mining of such assignments can be common un-confided in parties, contenders; in this way securing protection turns into a significant concern. Cryptographic strategies are unmistakably implied for such situations where numerous gatherings team up to process results or offer non-touchy mining results and subsequently keeping away from divulgence of delicate data. Cryptographic procedures locate its utility in such situations in light of two reasons: First, it offers an all-around characterized model for security that incorporates strategies for demonstrating and measuring it. Second a huge arrangement of cryptographic calculations and builds to actualize protection safeguarding information mining calculations are accessible in this space. The information might be conveyed among various colleagues vertically or evenly. In vertically apportioned information among various teammates, the individual elements may have various credits of same arrangement of records and in the event of on a level plane divided information, singular records are spread out over different elements, every one of which has similar arrangement of traits. The vast majority of the protection saving disseminated information mining calculations uncover nothing other than the conclusive outcome. Kantarcioglu and Clifton fused cryptographic strategies to protect security in affiliation rule mining over evenly apportioned information to limit data shared and simultaneously adding next to no overheads to the mining task. Lindell and Pinkas have examined how to produce ID3 choice trees on a level plane parcelled information. Yang et al. in have examined an answer for evenly divided information where every client has a private access just to their own record. Vaidya and Clifton were the main who concentrated how secure affiliation rule digging should be possible for vertically parcelled information. Du and Zhan introduced an answer for developing ID3 on vertically parcelled information contemplating two gatherings for mining. Vaidya and Clifton built up a Naive Bayes classifier for saving protection on vertically parcelled information. Vaidya and Clifton in proposed a technique for grouping over vertically apportioned information. Every one of these strategies are nearly founded on an exceptional encryption convention known as Secure Multiparty Computation (SMC) innovation. Yao in examined an issue where two tycoons needed to realize who is more extravagant with neither uncovering their total assets. Along these lines, SMC was begat and created. SMC characterizes two fundamental ill-disposed models specifically (I) Semi-Honest model and (ii) Malicious model. Semi-genuine model follows conventions truly, however can attempt to deduce the mystery data of different gatherings. In Malicious model, malevolent enemies can successfully construe mystery data. They can prematurely end convention, send false messages, conspire with different malevolent models or even satire messages. SMC utilized in circulated protection safeguarding information mining comprises of a lot of secure sub conventions that are utilized in on a level plane and vertically parcelled information: secure whole, secure set association, secure size of crossing point and scalar item. An itemized conversation of these conventions is given by Chris Clifton et al. in. Albeit cryptographic procedures guarantee that the changed information is precise and secure however this methodology neglects to convey when in excess of a couple of gatherings are included. Also, the information mining results may break the security of individual records. There exist a decent number of arrangements in the event of semi-legit models yet if there should be an occurrence of pernicious models less examination have been made.

Sharma, S., Chen, K., &Sheth, A. in their work developed a personalized health care information system for disease monitoring and they analysed their work and checked it how much privacy is preserved.

Zhang, C., Zhu, L., Xu, C., & Lu, R in their research work proposed a disease prediction system where there is a preservation of privacy where the medical data will be encrypted and the data is stored in the cloud server and it is stored for further processing like the training process by using the single layer perceptron learning model.

Dwivedi, A. D., Srivastava, G., Dhar, S., & Singh, R. in their work proposed the use of blockchain for securely managing and analysing the healthcare data. But the utilization of blockchain is costly and requires more bandwidth and high cost for computation and it is not suitable in the smart cities where the resource is constraint and they tried to resolve these issues in their work.

Zhou, J., Cao, Z., Dong, X., & Lin, X. in their work proposed a privacy preserving homomorphic and they developed a privacy preserving function correlation matching from the medical text mining and they found that their model achieved high security and good performance.

CONCLUSION

Protection of data identified with the information proprietor is the prime inspiration in the privacy preserving information mining measure. In the examination of protection safeguarding approaches it was discovered that the accessible calculations are not productively safeguarding the security at versatile framework. The shrouded data, be that as it may, can in any case be induced despite the fact that with some vulnerability level. An algorithm for sanitization calculation at that point could be estimated based on the vulnerability that it builds up during the remaking of the concealed data. Henceforth more exploration is important to guarantee the level of security also, capacity of ways to deal with diminish hidden failure data. Data mining for privacy preservation(PPDM) can be considered as promising field for secure and productive information concealing methodologies. It was recognized that productive grouping methodology is essential for concealing the information proprietor identity. Mechanism for classification will be utilized to diminish the dataset to veiled data which can be later utilized for the recovery of data. Regardless of above examined open examination issues, we accept that PPDM will be one of the most encouraging advancements for cutting edge information investigation. Hence this survey work will give some idea for the readers about the utilization of various algorithms in privacy preservation and how it can be applied to various domains.

References

1. Huang, Z., & Du, W. (2008, April). OptRR: Optimizing randomized response schemes for privacy-preserving data mining. In *2008 IEEE 24th International Conference on Data Engineering* (pp. 705-714). IEEE.
2. Zhu, Y., & Liu, L. (2004, August). Optimal randomization for privacy preserving data mining. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 761-766).
3. Erlingsson, Ú., Pihur, V., &Korolova, A. (2014, November). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (pp. 1054-1067).
4. Lohiya, S., &Ragha, L. (2012, November). Privacy preserving in data mining using hybrid approach. In *2012 Fourth International Conference on Computational Intelligence and Communication Networks* (pp. 743-746). IEEE.
5. Lin, J. L., & Liu, J. Y. C. (2007, March). Privacy preserving item set mining through fake transactions. In *Proceedings of the 2007 ACM symposium on Applied computing* (pp. 375-379).
6. Sharma, S., Chen, K., &Sheth, A. (2018). Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. *IEEE Internet Computing*, *22*(2), 42-51.
7. MS Kumar, KG Chandra, KK Gupta, "Stocks Analysis and Prediction of Indian Oil Trading Using Big Data Analytics", International Journal of Mechanical Engineering, Vol. 7 No. 1 January, 2022 pp: 6734-6738
8. *Anitha, C., M. Padmavathamma, and M. Sunil Kumar. "An Appraisal on privacy preserving mining of association rules." International Journal of Computer Applications 975: 8887.*
9. Kumar, M. S., & Harika, A. (2020). Extraction and classification of Non-Functional Requirements from Text Files: A Supervised Learning Approach. *Psychology and Education*, *57*(9), 4120-4123.
10. Pika, A., Wynn, M. T., Budiono, S., ter Hofstede, A. H., van der Aalst, W. M., &Reijers, H. A. (2019, September). Towards privacy-preserving process mining in healthcare. In *International Conference on Business Process Management* (pp. 483-495). Springer, Cham.
11. Pika, A., Wynn, M. T., Budiono, S., Ter Hofstede, A. H., van der Aalst, W. M., &Reijers, H. A. (2020). Privacy-preserving process mining in healthcare. *International journal of environmental research and public health*, *17*(5), 1612.

12. Evfimievski, A., & Grandison, T. (2009). Privacy-preserving data mining. In *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 527-536). IGI Global.
13. Rupesh, B., and M. Sunil Kumar. "Predicting the Hard Keyword Queries over Relational Databases." International Journal of Applied Engineering Research 10, no. 10 (2015): 26629-26640.
14. Qi, X., Mei, G., Cuomo, S., & Xiao, L. (2020). A network-based method with privacy-preserving for identifying influential providers in large healthcare service systems. *Future Generation Computer Systems*.
15. Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 439-450).
16. Kantarcioglu, M., & Clifton, C. (2004). Privacy-preserving distributed mining of association rules on horizontally partitioned data. *IEEE transactions on knowledge and data engineering*, *16*(9), 1026-1037.
17. Vaidya, J., & Clifton, C. (2003, August). Privacy-preserving k-means clustering over vertically partitioned data. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 206-215).
18. Vaidya, J., & Clifton, C. (2004, April). Privacy preserving naive bayes classifier for vertically partitioned data. In *Proceedings of the 2004 SIAM international conference on data mining* (pp. 522-526). Society for Industrial and Applied Mathematics.
19. Sangamithra, B., Swamy, B.M. and Kumar, M.S., 2021. A comparative study on a privacy protection in personalized web search. Materials Today: Proceedings.
20. Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557-570.
21. Madhan Subramaniam, S. R. (2010). An analysis on preservation of Privacy in Data Mining.
22. Bayardo, R. J., & Agrawal, R. (2005, April). Data privacy through optimal k-anonymization. In *21st International conference on data engineering (ICDE'05)* (pp. 217-228). IEEE.
23. Fung, B. C., Wang, K., & Yu, P. S. (2005, April). Top-down specialization for information and privacy preservation. In *21st international conference on data engineering (ICDE'05)* (pp. 205-216). IEEE.
24. Du, W., & Zhan, Z. (2003, August). Using randomized response techniques for privacy-preserving data mining. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 505-510).