NEGATIVE ASSOCIATION RULES FOR PROCESS MINING IN HEALTHCARE BASED ON PRIVACY PRESERVATION

Sameera Fatima [1], Md. Ateeq Ur Rahman[2], Jothikumar. R[3]

[1]Research Scholar, Department of Information Technology, SCET, Hyderabad, Telangana
Email: zany.sameera17@gmail.com

[2]Professor, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.
Email: mail_to_ateeq@yahoo.com

[3]Professor, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.
Email: drjothikumarr@gmail.com

ABSTRACT: Edge-distributed computing is an effective way to deal with address the high dormancy issue in portable distributed computing for administration provisioning, by setting a few figuring assets near end gadgets. To further develop the client fulfillment and the asset productivity, this paper centers around the undertaking offloading and administration reserving issue for offering types of assistance by edge-distributed computing. This paper figures out the issue as a compelled discrete streamlining issue, and proposes a crossover heuristic technique in view of Molecule Swarm Advancement (PSO) to tackle the issue in polynomial time. The proposed technique, LMPSO, exploit PSO to take care of the assistance reserving issue. To keep away from PSO catching into nearby streamlining, LMPSO adds lévyflight development for molecule refreshing to work on the variety of molecule. Given the help reserving arrangement, LMPSO involves a heuristic strategy with three phases for task offloading, where the main stage attempts to make full utilization of cloud assets, the subsequent stage involves edge assets for fulfilling necessities of inactivity touchy errands, and the last stage works on the general execution of undertaking executions by re-offloaded a few assignments from the cloud to edges. Reenacted analyze results show that LMPSO has up to 156% better client fulfillment, up to 57.9% higher asset effectiveness, and up to 155% more noteworthy handling proficiency, in general, thought about with other seven heuristic and meta-heuristic strategies.

I. INTRODUCTION

Electronic flourishing records (EHRs) are in general elaborate by a grouping of clinical benefits relationship with an outrageous objective to work on understanding thought and update the efficiency of clinical benefits transport. In complex clinical circumstances, the EHR structure speeds up the clinician's work cycle through motorizing the data the trailblazer's cycle. When utilized really, these EHRs work with various standard clinical benefits endeavors, yet next to help in the particular clear check of weights. Individuals' approval to their clinical records is worked with by EHRs. Basically, they go with a home achievement seeing plan that grants patients to reliably measure and evaluate their unintentional effects. The spread of data from the EMR structure is central for dealing with the chance of clinical assessment. Experts use this data to play out countless endeavors including data mining, similar to portrayal (doubt for diabetic presence), gathering (risk unquestionable check), genuine tests (weight record and diabetes affiliation), or requesting replying.

As well as reviving the finished human relationship to patients, experts in clinical benefits should benefit from the joining of information and electronic achievement records (EHRs). Fig.1 shows the key data mining applications in clinical idea; see for extra nuances. Anonymization-based procedures and cryptographic structures have both been presented in the piece as framework for achieving security. Anonymization is thoroughly used by experts inferable from the lower correspondence and computation costs of indistinct from appeared particularly similar to their cryptographic colleagues. Information mishap is one of the fundamental issues in an anonymization-based approach. The information event sorts out the detachment between the fundamental illuminating records and the anonymized data bases. The information catastrophe increases with the development in the level of the hypothesis as well as mask framework.

All around, the information disaster should be less to achieve higher data utility. K-secret is a key felt that had a ton of familiarity with address the bet of re-seeing confirmation of anonymized data through linkage to other datasets. K-Mystery can hinder character divulgence, i.e., a record in the k-

anonymized enlightening report can't be set up back to the relating record in the fundamental educational overview. At any rate, as a last resort, it could forget to defend against trademark openness. A few intellectual individuals of k-anonymization can't fight the temptation to conflict with the way that achieving a re-ID peril of zero is odd or boundless. By a wide margin the vast majority of approaches fall into two general classes: those that safeguard sensitive data during mining and those that protect fragile data mining yields. The major blueprint consolidates procedures like weight, inspecting, and change to pass cleaned datasets that may be safely given on to various get-togethers. These techniques are expected to help data diggers with stop by immense results anyway, when they have no obvious data. Of course, the accompanying class consolidates systems for keeping sensitive data plans got from the use of data mining appraisals enigmatic, looking into methods for cleaving down the presentation of classifiers for portrayal endeavors so delicate information isn't revealed. With the approaching of monstrous data mining gadgets and plans, we are stood up to with the primer of building a clinical idea data mining system that meets client hypotheses for fundamental data revelation from instructive combinations while avoiding the likelihood to gather individual information about people. An individual or association should not be fit to be seen (by an outcast) considering the mining methods or results that we share. Getting fragile information far from being passed on in an unapproved or unconstrained way is the goal of Affirmation Saving Data Mining (PPDM). In clinical idea systems, the PPDM technique is completely useful for limiting the spillage of sensitive individual information. Likewise, it engages the evaluation of gigantic enlightening assortments from clinical assessment structures to get information and see consistent strategies for perilous weights without chasing after patients' security. Regardless of what the way that enlightening grouping affiliation structures have constantly centered around data security, mining data and keeping the responsiveness of shaky information has now changed into the most crushing and most raised need objective of the data mining process.

Security defending methodologies endeavor to avoid the spillage of fragile data, making it more testing to track down sensitive data from non-fragile data. Notwithstanding, they don't pardon the exposure of acceptance rules. Hence, experts determinedly stand isolated on insurance saving organization controls lately. Affiliation rule structures have been comprehensively used in an enormous mix of attempts and current firms across a wide combination of industry regions, including advancing, estimating, diagnostics, and security. Sensitive affiliation rule covering is a subsection of PPDM that falls under the class of result insurance. Restrictive standards are those that should be kept secret. One of the PPDM methodology used to keep stowed away information hidden away is known as data decontaminating. Instructive combination purging computations that shield fragile nonproduction illuminating variety data from unapproved access are being made and had a go at using a level of approaches at this point open in the association. These evaluations fall under the going with plans:

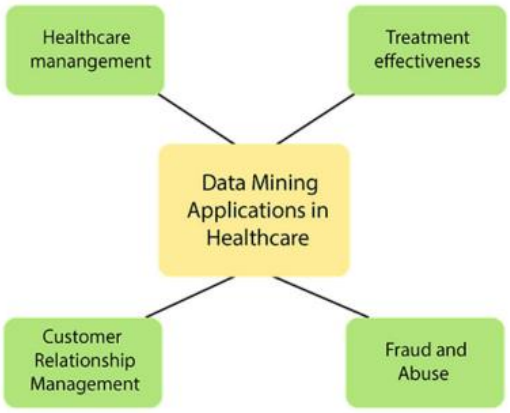
(1) Appraisals control the covering structure using the standard's assistance or sureness; and

(2) Appraisals change crude data by harming or diverting the fundamental characteristics. It is possible that the change cycle changes the crucial plan of assumes that may be mined from the fundamental illuminating grouping, either by disguising non-sensitive standards (lost rules) or by introducing induces that were not stayed aware of by the essential data base into the mining framework (indistinct vision rules). We endeavored to work with these terrible effects by changing the first dataset a piece and fittingly. Considering the difficulty of PPDM, which is known as a NP-bothersome issue, finding the best outline with the most un-inadvertent impacts is hard.

Using negative affiliation rules to find covering discretionary effects or drugs that limit strikingly together is a useful gadget in clinical idea data assessment. Negative affiliation rule mining is endeavoring to accomplish considering the basic capacities among positive and negative association rule mining. Concerning mining negative alliance rules, there are two key concerns that ought to be overseen by the researchers: looking for and filtering the negative affiliation rules. An Inherited Computation (GA) looks for a response using a general public of centers rather than a lone point. This approach is both computationally speedy and fruitful. Fabulous Pursuit (TS) sees at each string as a point in the game plan space.

TS leads cycles beginning with one area point then onto the join by invigorating the approaches locally and can avoid starved close by minima. Joining GA and TS with their different resources manages the likelihood of finding a sensible solution for overall combinatorial update issues. GA begins with different preliminary plans and makes a lot of creative plans using the cross grouping search approach. TS fosters every plan of stand-separated methodologies through

a close by seek after. GA then, at that point, consolidates the extended plan of TS to keep a close to improvement.



A. MOTIVATION

Clinical advantages process information can merge interminable delicate variables and altogether clashing cycle ways to deal with acting that current additional protection issues. The clinical thought industry should con shape to outrageous information security rules. Security insurance for such information while remaining mindful of its handiness for process mining is a consistent worry in clinical thought. There is a compromised between information security and worth in the utilization of existing information change techniques for anonymizing clinical advantages process information. For instance, encryption doesn't give sufficient protection security when used to upgrade the worth of information for process mining. The accuracy of results might be compromised if methodology that adjustment of agreement with more serious affirmation rules, (for example, speculation) are utilized. Recorded as a printed version, immense examination has utilized secret, information covering, information unsettling influence, and cryptography for information protection. Utilizing dynamic information covering, we can accomplish arrangement safeguarding veiling and anonymization without having to genuinely duplicate information or kill values — errands which can concede appraisal, yet can disable the utility of information and present the bet of human slip up. The cryptographic. Approach is particularly challenging to scale when more than a barely any social gatherings are involved. It in addition doesn't resolve whether or not uncovering the last information mining results might disregard the security of individual records. The disturbance approach doesn't repeat the essential information values. New assessments have been made to recreate the fundamental information scattering. All around, every system has its own horrendous engravings, i.e., data difficulty, security break, and ignorant

Utility. It is not difficult to see that secret isn't satisfactory. For instance, acknowledge we utilize k-absence of clearness to safeguard information. This really designs that, knowing seeing data about a person, there are essentially k records in the educational file that could (with indistinguishable likelihood) suggest that person. Regardless, acknowledge that those records moreover merge delicate data, e.g., it is diabetic to expect to be a person. Expecting that all k people have a similar worth for interesting data (for instance, all are diabetic), then, at that point, k-secret offers no insurance against receptiveness of that reality. This has incited substitute techniques. In any case, it is at this point testing to answer the solicitation, "is the information perplexing enough?". Various sophistications of k-secret have been recommended, similar to p-delicate k-secret, l-combination, and t-closeness. An edifying document is said to fulfill l-variety expecting there are basically l for the most part around tended to values for each classified quality in every get-together of records that arrangement key credits. Notwithstanding, this improvement experiences a similitude assault. If the expected additions of a sensitive quality in a party are l-different yet the semantics are something essentially the equivalent, the trademark will comparatively be revealed. Nonetheless, the work in [8] handles this issue by proposing semantic combination, which develops l-collection. See for extra subtleties. As to, delicate k-namelessness, its motivation is to shield against brand name disclosure by expecting that there be in a general sense p various qualities for each strange quality inside the records sharing a mix of key credits. P-delicate k-obscurity has the deterrent of most likely expecting that each baffling property takes esteems consistently over its space, at the end of the day, that the frequencies of the different likely gains of an arranged brand name are comparative. Whenever this isn't exact, accomplishing p-delicate k-absence of clearness could cause an immense information utility misfortune. T-closeness watches out for the property receptiveness inadequacies regular to l-grouping: Skewness assault, since inside pack development of private credits is indistinguishable from the vehicle of those ascribes for the whole dataset, no skewness assault can happen. Once more comparability assault, since inside pack stream of private credits copies the dissipating of those ascribes over

The whole dataset, no semantic similarity can happen inside a pack that doesn't happen in the whole dataset. Obviously, inside pack equivalence can't be abstained

from accepting all patients in an illuminating arrangement have comparative diseases. Information digging systems for security assurance can be general or express. Information mining tasks could utilize normal approaches to overseeing change information into a plan that can be utilized as a data. Changing records without adding new respects or changing existing qualities might be used to accomplish absence of lucidity utilizing these strategies (e.g., information trading) (e.g., by adding disturbance). Certain information mining strategies remember pri vacy affirmations for their assessments (e.g., protection guarding choice tree demand). Anonymization procedures, (for instance, association rule stowing away) have been portrayed for touchy information mining yields as well.**B. OUR CONTRIBUTION**

The following is a summary of the article's significant contributions:

1) The main and important methods for data security through safe sharing in a cloud context are reviewed in this paper.

2) We offer the following information regarding each strategy.

(a)How it functions in terms of data protection, and (b) the superior, ground-breaking options available.

In order to make it simple for readers to understand the essence of the approach as well as its applications, we also include potential and valuable information about each presented solution in a tabular manner, such as its working, implementation environment, success, range of the provided model, etc.

3) A thorough and comparative study of the methodologies covered is conducted and presented in an accessible manner. Additionally, research is done to determine which technique will best meet the needs.

II.EXISTING SYSTEM:

Defending the security of clinical consideration information is a critical piece of enabling data supervisors to give exact records so that mining could go on with conviction. The usage of connection rule mining in clinical benefits data has been all over to this second. Most applications place on specific connection rules, disregarding the unfavorable consequences of explicit scientific strategies. With respect to interfacing divergent contaminations and drugs, negative connection rules could give more steady information than positive ones. This is especially clear with respect to specialists and social affiliations (e.g., a particular aftereffect will not arise when certain incidental effects exist). Data mining in clinical benefits ought to be done such that defends the character of patients, especially while overseeing sensitive information. Nevertheless, uncovering this information imperils it of attack. Clinical consideration data security confirmation has of late been tended to by headways that irritated data (data disinfection) and reproduce all out dispersals considering a genuine worry for doing investigate in data mining.

Existing System Disadvantages:

- Data has been widespread to this point in time.
- When dealing with sensitive information.
- Data privacy protection has lately been addressed by technologies that disrupt data.

III.LITERATURE SURVEY

TITLE: Analyzing and performing privacy preserving data mining on medical databases.

DESCRIPTION:

For both the production and consumption of data the internet is becoming a standard whereas the security for private data is gradually decreasing. Therefore, to have a safe transaction in the data, security and privacy would be the key issues to be considered. In recent days, privacy has become a key issue in many data mining and knowledge discovery fields which lead to the development of many Privacy Preserving Data Mining (PPDM) techniques. In our work we use few of these techniques to privately preserve the data holder such as hospital data. In this we use techniques named “Anonymization”, “Suppression”, “Generalization” and “Data Hiding” on different fields for the data to be more secure and project the data which is useful to the public. This is a new way of our approach to create awareness among the public to be more attentive and health conscious. The modified data is clustered based on diseases. Based on the end user requirement the private data of the individual is hidden and the required data is projected.

TITLE: Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data.

DESCRIPTION:

The trends of data mining in the healthcare is increased due to the digitization of healthcare with electronic health record (EHR) systems. This generates a huge amount of data on daily basis. Data mining with the healthcare data has given the new direction to medical research for early detection of diseases and improving

patient care. Many data mining applications require the integration of data from the different sources. For example, the integration of outpatient medical records and health examination data helps to identify the correlation between abnormal test result and disease. The result of association rule mining on this integrated data helps to build the knowledge center for disease prevention, which facilitate the healthcare provider in follow up treatment and prevention. The integration of data requires the sharing of sensitive information about the patients. Disclosing the sensitive information violates the privacy of patients. In this paper, we tackle the problem of privacy preserving association rule mining in vertically partition healthcare data. Furthermore, we analyze the proposed approach in terms of privacy preservation, communication and computation cost.

TITLE: Privacy preserving data mining on published data in healthcare: A survey.
DESCRIPTION:
Healthcare data is considered very significant to researchers in this field. Such information must be published with methods that keep the identity of patients hidden especially when dealing with sensitive information. Publishing such information makes it more vulnerable to attackers. As such, many techniques were proposed to preserve the privacy of healthcare data. In this paper, we illustrated a survey for the models and techniques that are used for publishing data about patients.

TITLE: Towards privacy-preserving process mining in healthcare.

DESCRIPTION:
Process mining has been successfully applied in the healthcare domain and helped to uncover various insights for improving healthcare processes. While benefits of process mining are widely acknowledged, many people rightfully have concerns about irresponsible use of personal data. Healthcare information systems contain highly sensitive information and healthcare regulations often require protection of privacy of such data. The need to comply with strict privacy requirements may result in a decreased data utility for analysis. Although, until recently, data privacy issues did not get much attention in the process mining community, several privacy-preserving data transformation techniques have been proposed in the data mining community. Many similarities between data mining and process mining exist, but there are key differences that make privacy-preserving data mining techniques unsuitable to anonymize process data. In this article, we analyze data privacy and utility requirements for healthcare process data and assess the suitability of privacy-preserving data transformation methods to anonymize healthcare data. We also propose a framework for privacy-preserving process mining that can support healthcare process mining analyses.

TITLE: Sharing healthcare information based on privacy preservation.
DESCRIPTION:
The evolution and development of information technology have facilitated greater sharing of data and knowledge management for the collection of electronic information by data owners such as governments, corporations, and individuals. Therefore, they have created huge opportunities for knowledge management and information retrieval. Recent developments have helped improve decision making especially in the fields of medical information, research, and public health organization, among others. Recently, the control and sharing of data or knowledge management has received notable attention in research communities. Many approaches have been proposed for different data publishing needs in different fields. The sharing of data needs control and management to ensure system integration. Integration is required especially in the management of patient data to secure sensitive information such as patient identification. Several studies have focused on the management of data in medical applications to ensure system integration. However, the management and sharing of data in different fields may result in misuse of information. Therefore, there is a need to build models or design certain algorithms to manage shared data efficiently and to avoid misuse. The goal is to ensure authenticity of the data system. In the present study, we systematically summarize and evaluate different approaches to control the sharing of data and knowledge management in order to ensure system integration. Moreover, we study the challenges in controlling the sharing of data and clarify the differences and conditions that distinguish the control of sharing of data from other related problems. Finally, we correspondingly propose future research directions in the conclusion.

TITLE: Privacy preserving data mining for healthcare record: A survey of algorithms.
DESCRIPTION:
If one is to believe the popular press and many “technical writings,” blockchains create not only a perfect transactional environment but also obviate the need for banks, lawyers and courts. The latter will soon be replaced by smart contracts: unbiased and infallible computer programs that form, perform and enforce agreements. Predictions of future revolutions must,

however, be distinguished from the harsh reality of the commercial marketplace and the technical limitations of blockchain technologies. The fact that a technological solution is innovative and elegant need not imply that it is commercially useful or legally viable. Apart from attempting a terminological "clean-up" surrounding the term smart contract, this paper presents some technological and legal constraints on their use. It confronts the commonly made claims concerning their ability to automate the transacting process and to ensure perfect performance. It also examines the possibility of reducing contractual relationships into code and the ability to integrate smart contracts with the complexities of the real world. A closer analysis reveals that smart contracts can hardly be regarded as a semi-mythical technology liberating the contracting parties from the shackles of traditional legal and financial institutions. While some of their technical features seem prima facie attractive, especially to non-lawyers, a closer analysis reveals their many shortcomings.

IV.PROPOSED SYSTEM

In this review, metaheuristic-based information disinfection for medical care information mining is explored to keep patient security safeguarded. It is trusted that by involving the Forbidden hereditary calculation as a streamlining apparatus, the proposed method picks thing sets to be disinfected (adjusted) from exchanges that fulfill delicate negative measures fully intent on limiting changes to the first data set. Explores different avenues regarding benchmark medical care datasets show that the recommended protection safeguarding information mining (PPDM) strategy beats existing calculations concerning Concealing Disappointment (HF), Counterfeit Rule Age (AR), and Lost Rules (LR).

PROPOSED SYSTEM ADVANTAGE

- Providing more security
- Data mining is investigated in order to keep patient privacy protected.
- Transactions that satisfy sensitive negative criteria with the goal of minimizing changes to the original database.

V.MODULES NAME

1. User

2. Item Extractions

3. Negative Rules

4. Genetic Algorithm

1. User

In this module we design the windows for the project. These windows are used for secure login for all users. To connect with server user must give their username and password then only they can able to connect the server. If the user already exits directly can login into the server else user must register their details such as username, password and Email id, into the server. Server will create the account for the entire user to maintain upload and download rate. Name will be set as user id. Logging in is usually used to enter a specific page.

2. Item Extractions

This is the first module Data User can register and Login. After login Data User have an option of searching the files as a file name. Data user can also have a download file it will show an encrypted data. Data user can also send a trapdoor request to the server. Server can accept the request and then data user can takes permissions from the owner then the file it will downloaded in plain text

3. Negative Rules

This is the Second module of this project. In this module Data Owner should register and Login. Data Owner will Uploads the files into the database. Data owner can also send request to the data user.

4. Genetic Algorithms

This is the third module of this project. In this module Cloud Server can login. After login it will see all data owners' information. Cloud server can see all users' information. Cloud server can see an all stored data files. Cloud server can give keys request to the user. Cloud server can also see an attacker information of file.

VI. Proposed Algorithm

Proposed Algorithm

Linear programming (LP) problems to obtain the encrypted MWTC

Treating the matching between queries and documents as an optimal matching task, we formulate the word transportation (WT) problem following the optimal transportation problem of linear programming. We utilize WT problems to calculate the minimum word

transportation cost (MWTC) as the similarity metric between queries and documents.

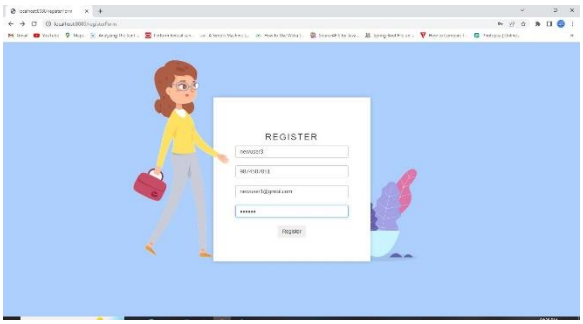
We introduce the forward indexes as semantic information of documents. We define each keyword and its weight in the forward index of a document as the keywords distributions for the document. Therefore, we need to select keywords for each document and calculate the weight of each keyword in a specific document. Without loss of generality, we use TF-IDF (term frequency inverse document frequency) as a criterion to select keywords in our scheme.

VII.EXISTING SYSTEM
kNN algorithm

In introduced homomorphic encryption to encrypt relevance scores and realize a multi keyword ranked search scheme under the vector space model. Encryption techniques. It is not supporting multi-keyword ranked searching schemes that can resist against several attacks brought by OPE-based schemes. Secure Semantic Searching.

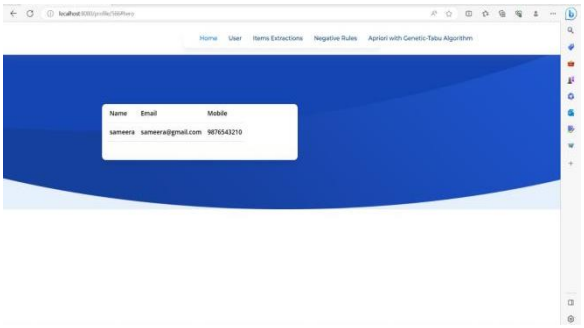
VIII.SCREENSHOTS

Register Page

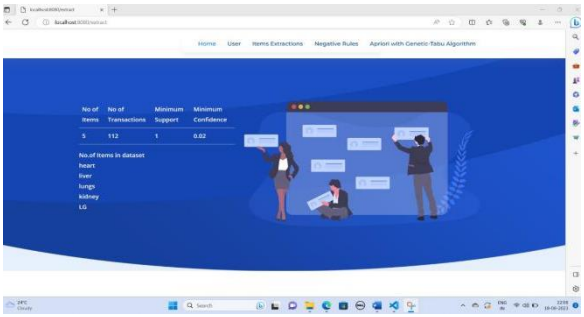


User Home Page

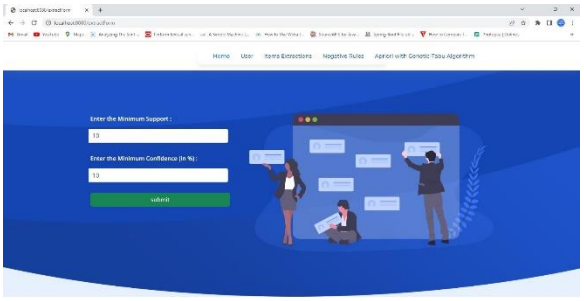
User Profile Page



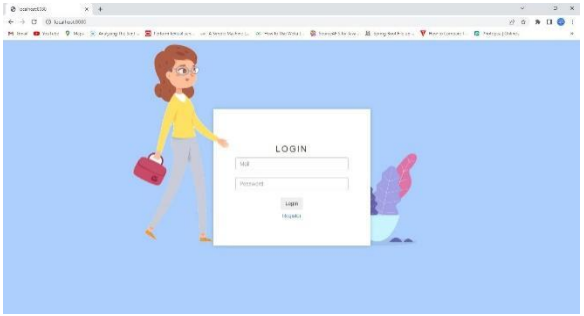
Item Extraction page



A prior Algorithm Page



Login Page



IX. FUTURE ENHANCEMENT

Protection safeguarding information change methods, log expansions, and cycle mining calculations will be generally inspected in future work, as well as an exact examination of what these techniques mean for

medical care logs. Moreover, the recommended strategy will be tried on an assortment of medical services datasets, incorporating those with fluctuating highlights, to guarantee that it is powerful. For more protection, negative affiliation rules will be applied with differential security, neighborhood differential security, or a mix of both.

X. CONCLUSION

Keeping and moving clinical information has become more troublesome as worries about protection have developed. The spread of medical services information might be extremely helpful, yet it should be finished in a way that safeguards the protection of patients. It's anything but a simple errand to guarantee the protection of the information that has been dispersed. The medical services industry has a predicament in safeguarding patient information while likewise making it helpful for information mining. There are various security safeguarding information mining techniques being used today. How much protection presented by every one of these calculations might be grouped in one of three ways: strategy based security, factual protection, or a blend of the three. Thus, protection and information use rules for various types of medical care information might be incredibly changed. The reason for this study is to address the protection issues related with medical services data sets because of information mining advancements. We utilized a hereditary enhancement procedure to conceal negative delicate affiliation rules utilizing a heuristic methodology in view of both mutilation and limitation processes. The proposed arrangement depends on a methodology of simultaneously diminishing the trust in the touchy principles. The procedure makes the least potential adjustments to the information base and misses the least conceivable non-delicate affiliation rules, which is a definitive objective of information sterilization. The proposed calculation is a crossover of the Apriori and incorporated hereditary Forbidden calculations. As opposed to mining negative affiliation governs instinctively, the proposed procedure uses negative intriguing quality to portray and make sense of the outcome of negative affiliation rules. By utilizing a hereditary Unthinkable hunt technique, the framework brings down the mining system's inquiry space. The fundamental advantages of the calculation are that

(1) A straightforward heuristic technique is utilized to pick the exchanges and things to be cleaned;

(2) A hereditary calculation is utilized to change the casualty's selection of things; and

(3) Information accessibility is improved by concealing principles rather than things. The wellness capability's productivity has been assessed in an assortment of medical care data sets to decide if it holds up when different changes are made to the first data set. From the reproduction results, obviously the guidelines of the recommended method have a lot higher help and certainty values while requiring substantially less handling time to arrive at the objectives.

XI.REFERENCES

[1] D. A. Kumari, Y. Vineela, T. M. Krishna, and B. S. Kumar, ‘‘Analyzing and performing privacy preserving data mining on medical databases,’’ Indian J. Sci. Technol., vol. 9, no. 17, pp. 1–9, May 2016.
[2] N. Domadiya and U. P. Rao, ‘‘Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data,’’ Proc. Comput. Sci., vol. 148, pp. 303–312, Jan. 2019.
[3] L. A. Abuwardih, W. Shatnawi, and A. Aleroud, ‘‘Privacy preserving data mining on published data in healthcare: A survey,’’ in Proc. 7th Int. Conf. Comput. Sci. Inf. Technol. (CSIT), Jul. 2016, pp. 1–6.
[4] A. Pika, T. Wynn, S. Budiono, A. Hofstede, W. Aalst, and H. Reijers, ‘‘towards privacy-preserving process mining in healthcare,’’ in Proc. Int. Conf. Bus. Process Manage. 2019, pp. 483–495.
[5] H. R. Asmaa and B. M. Y. Norizan, ‘‘Sharing healthcare information based on privacy preservation,’’ Sci. Res. Essays, vol. 10, no. 5, pp. 184–195, Mar. 2015.
[6] M. Hassan, M. A. Butt, M. Zaman, and U. of Kashmir, ‘‘Privacy preserving data mining for healthcare record: A survey of algorithms,’’ Int. J. Trend Sci. Res. Develop., vols. 2, no. 1, pp. 1176–1184, Dec. 2017.
[7] A. Pika, T. Wynn, and S. Budiono, ‘‘Privacy-preserving process mining in healthcare,’’ Int. J. Environ. Res. Public Health, vol. 17, no. 5, pp. 1–28, 2020.
[8] K. Oishi, Y. Sei, Y. Tahara, and A. Ohsuga, ‘‘Semantic diversity: Privacy considering distance between values of sensitive attribute,’’ Comput. Secur., vol. 94, Jul. 2020, Art. no. 101823.
[9] S. Darwish, M. Madbouly, and M. El-Hakeem, ‘‘A database sanitizing algorithm for hiding sensitive multi-level association rule mining,’’ Int. J. Comput. Commun. Eng., vol. 3, no. 4, p. 285, 2014.
[10] M. N. Dehkordi, K. Badie, and A. K. Zadeh, ‘‘A novel method for privacy preserving in association rule mining based on genetic algorithms,’’ J. Softw., vol. 4, no. 6, pp. 555–562, Aug. 2009.
[11] R. Crawford, M. Bishop, B. Bhumiratana, L. Clark, and K. Levitt, ‘‘Sanitization models and their

limitations,'' in Proc. Workshop New Secur. Paradigms, 2006, pp. 41–56.
[12] N. Suryawanshi, S. Jain, and A. Jain, ''Mining interesting positive and negative association rule based on improved genetic algorithm (MIPNAR_GA),'' Int. J. Adv. Comput. Sci. Appl., vol. 5, no. 1, pp. 7834–7845, 2014.
[13] J. C.-W. Lin, J. M.-T. Wu, P. Fournier-Viger, Y. Djenouri, C.-H. Chen, and Y. Zhang, ''A sanitization approach to secure shared data in an IoT environment,'' IEEE Access, vol. 7, pp. 25359–25368, 2019.
[14] D. Toshniwal, ''Privacy preserving data mining techniques for hiding sensitive data: A step towards open data,'' in Data Science Landscape (Studies in Big Data), vol. 38. Singapore: Springer, 2018, pp. 205–212.
[15] V. S. Verykios, E. Bertino, I. N. Fovino, L. P. Provenza, Y. Saygin, and Y. Theodoridis, ''State-of-the-art in privacy preserving data mining,'' ACM SIGMOD Rec., vol. 33, no. 1, pp. 50–57, 2004.
[16] E. Dasseni, V. Verykios, A. Elmagarmid, and E. Bertino, ''Hiding association rules by using confidence and support,'' in Proc. Int. Workshop Inf. Hiding, 2001, pp. 369–383.
[17] R. Oliveira and O. Zaiane, ''Privacy preserving frequent itemset mining,'' in Proc. IEEE Int. Conf. Privacy, Secure Data Mining, Dec. 2002, pp. 43–54.
[18] M. Z. Islam and L. Brankovic, ''Privacy preserving data mining: A noise addition framework using a novel clustering technique,'' Knowl.-Based Syst., vol. 24, no. 8, pp. 1214–1223, 2011.
[19] X. Liu, S. Wen, and W. Zuo, ''Effective sanitization approaches to protect sensitive knowledge in high-utility itemset mining,'' Int. J. Speech Technol., vol. 50, no. 1, pp. 169–191, Jan. 2020.
[20] X. Sun and P. S. Yu, ''Hiding sensitive frequent itemsets by a border-based approach,'' J. Comput. Sci. Eng., vol. 1, no. 1, pp. 74–94, Sep. 2007.
[21] G. V. Moustakides and V. S. Verykios, ''A MaxMin approach for hiding frequent itemsets,'' Data Knowl. Eng., vol. 65, no. 1, pp. 75–89, Apr. 2008.
[22] A. Amiri, ''Dare to share: Protecting sensitive knowledge with data sanitization,'' Decis. Support Sys., vol. 43, no. 1, pp. 181–191, 2007.
[23] S.-L. Wang, B. Parikh, and A. Jafari, ''Hiding informative association rule sets,'' Expert Syst. Appl., vol. 33, no. 2, pp. 316–323, Aug. 2007.
[24] Y. H. Wu, C. M. Chiang, and A. L. P. Chen, ''Hiding sensitive association rules with limited side effects,'' IEEE Trans. Knowl. Data Eng., vol. 19, no. 1, pp. 29–42, Jan. 2007.
[25] A. Gkoulalas-Divanis and V. S. Verykios, ''Exact knowledge hiding through database extension,'' IEEE Trans. Knowl. Data Eng., vol. 21, no. 5, pp. 699–713, May 2009.
[26] C.-M. Wu and Y.-F. Huang, ''A cost-efficient and versatile sanitizing algorithm by using a greedy approach,'' Soft Comput., vol. 15, no. 5, pp. 939–952, May 2011.
[27] P. Cheng, I. Lee, J.-S. Pan, C.-W. Lin, and J. F. Roddick, ''Hide association rules with fewer side effects,'' IEICE Trans. Inf. Syst., vol. E98.D, no. 10, pp. 1788–1798, 2015.
[28] T.-P. Hong, C.-W. Lin, K.-T. Yang, and S.-L. Wang, ''Using TF-IDF to hide sensitive itemsets,'' Appl. Intell., vol. 38, no. 4, pp. 502–510, 2012.
[29] C. Wei, T. Hong, J. Wong, G. Lan, and W. Lin, ''A GA-based approach to hide sensitive high utility itemsets,'' Sci. World J., vol. 2014, Jan. 2014, Art. no. 804629.
[30] U. Yun and J. Kim, ''A fast perturbation algorithm using tree structure for privacy preserving utility mining,'' Expert Syst. Appl., vol. 42, no. 3, pp. 1149–1165, Feb. 2015.
[31] J. M.-T. Wu, J. Zhan, and J. C.-W. Lin, ''Ant colony system sanitization approach to hiding sensitive itemsets,'' IEEE Access, vol. 5, pp. 10024–10039, 2017.
[32] T. Wu, J. Lin, Y. Zhang, and C. Chen, ''A grid-based swarm intelligence algorithm for privacy-preserving data mining,'' Appl. Sci., vol. 9, no. 4, p. 774, 2019.
[33] A. Divanis and V. Verykios, ''An overview of privacy preserving data mining,'' ACM Mag. Students, vol. 15, no. 4, pp. 23–26, 2009.
[34] P. L. Lekshmy and M. A. Rahiman, ''A sanitization approach for privacy preserving data mining on social distributed environment,'' J. Ambient Intell. Humanized Comput., vol. 11, no. 7, pp. 2761–2777, Jul. 2020.
[35] J. M.-T. Wu, G. Srivastava, A. Jolfaei, P. Fournier-Viger, and J. C.-W. Lin, ''Hiding sensitive information in eHealth datasets,'' Future Gener. Comput. Syst., vol. 117, pp. 169–180, Apr. 2021.
[36] J. M. Wu, G. Srivastava, U. Yun, S. Tayeb, and J. C. Lin, ''An evolutionary computation-based privacy-preserving data mining model under a multithreshold constraint,'' Trans. Emerg. Telecommun. Technol., vol. 32, no. 3, Mar. 2021, Art. no. e4209.
[37] A. S. M. T. Hasan, Q. Jiang, J. Luo, C. Li, and L. Chen, ''An effective value swapping method for privacy preserving data publishing,'' Secur. Commun. Netw., vol. 9, no. 16, pp. 3219–3228, Nov. 2016.
[38] A. Zigomitros, F. Casino, A. Solanas, and C. Patsakis, ''A survey on privacy properties for data publishing of relational data,'' IEEE Access, vol. 8, pp. 51071–51099, 2020.
[39] K. Ranjith and A. G. Mary, ''Privacy-preserving data mining in spatiotemporal databases based on mining negative association rules,'' in Emerging Research in Data Engineering Systems and Computer

Communications (Advances in Intelligent Systems and Computing), vol. 1054. Singapore: Springer, 2020, pp. 329–339.
[40] A. Telikani and A. Shahbahrami, ‘‘Data sanitization in association rule mining: An analytical review,’’ Expert Syst. Appl., vol. 96, pp. 406–426, Apr. 2018.
[41] J. Lin, T. Wu, P. Fournier-Viger, G. Lin, T. Hong, and J. Pan, ‘‘A sanitization approach of privacy preserving utility mining,’’ in Proc. Int. Conf. Genetic Evol. Comput. Cham, Switzerland: Springer, 2015, pp. 47–57.
[42] S. Bagui and P. C. Dhar, ‘‘Positive and negative association rule mining in Hadoop’s MapReduce environment,’’ J. Big Data, vol. 6, no. 1, pp. 1–6, Dec. 2019.
[43] A. S. A. Kadir, A. A. Bakar, and A. R. Hamdan, ‘‘Frequent absence and presence itemset for negative association rule mining,’’ in Proc. 11th Int. Conf. Intell. Syst. Design Appl., Nov. 2011, pp. 965–970.
[44] C. Cornelis, P. Yan, X. Zhang, and G. Chen, ‘‘Mining positive and negative association rules from large databases,’’ in Proc. IEEE Conf. Cybern. Intell. Syst., Jun. 2006, pp. 1–6.
[45] N. Suryawanshi, S. Jain, and A. Jain, ‘‘Mining interesting positive and negative association rule based on improved genetic algorithm (MIPNAR_GA),’’ Int. J. Adv. Comput. Sci. Appl., vol. 5, no. 1, pp. 1–10, 2014.
[46] N. Rai, S. Jain, and A. Jain, ‘‘Mining positive and negative association rule from frequent and infrequent pattern based on IMLMS_GA,’’ Int. J. Comput. Appl., vol. 77, no. 14, pp. 48–52, Sep. 2013.
[47] S. Narmadha and S. Vijayarani, ‘‘Protecting sensitive association rules in privacy preserving data mining using genetic algorithms,’’ Int. J. Comput. Appl., vol. 33, no. 7, pp. 34–37, 2011.
[48] P. RajyaLakshmi, C. M. Rao, M. Dabbiru, and K. V. Kumar, ‘‘Sensitive itemset hiding in multi-level association rule mining,’’ Int. J. Comput. Sci. Inf. Technol., vol. 2, no. 5, pp. 2124–2126, 2011.
[49] F. Ullah, I. Ullah, A. Khan, M. I. Uddin, H. Alyami, and W. Alosaimi, ‘‘Enabling clustering for privacy-aware data dissemination based on medical healthcare-IoTs (MH-IoTs) for wireless body area network,’’ J. Healthcare Eng., vol. 2020, pp. 1–10, Nov. 2020.
[50] M. M. Madbouly, S. M. Darwish, N. A. Bagi, and M. A. Osman, ‘‘Clustering big data based on distributed fuzzy K-medoids: An application to geospatial informatics,’’ IEEE Access, vol. 10, pp. 20926–20936, 2022.
[51] U. Ahmed, G. Srivastava, and J. Lin, ‘‘A machine learning model for data sanitization,’’ Comput. Netw., vol. 189, Apr. 2021, Art. no. 107914.
[52] J. M.-T. Wu, G. Srivastava, M. Pirouz, and J. C.-W. Lin, ‘‘A GA-based data sanitization for hiding sensitive information with multi-thresholds constraint,’’ in Proc. Int. Conf. Pervasive Artif. Intell. (ICPAI), Dec. 2020, pp. 29–34.
[53] A. Khedr, W. Osamy, A. Salim, and A. Salem, ‘‘Privacy preserving data mining approach for IoT based WSN in smart city,’’ Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 8, pp. 555–563, 2019.
[54] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, ‘‘A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT,’’ IEEE Access, vol. 5, pp. 3302–3312, 2017.
[55] H. Li, F. Guo, W. Zhang, J. Wang, and J. Xing, ‘‘(a,k)-anonymous scheme for privacy-preserving data collection in IoT-based healthcare services systems,’’ J. Med. Syst., vol. 42, no. 3, pp. 1–9, Mar. 2018.