Preserving Sensitive Information using Fuzzy C-Means Approach

Asha Kiran Grandhi
Rajarshi Shahu College of Engineering
Pune
Maharashtra, India

Manimala Puri, PhD
Director,
JSPM Group of Institutions
Pune, Maharashtra, India

S. Srinivasa Suresh, PhD
Associate Professor,
CSE Department, KMIT,
Hyderabad, Telangana, India

ABSTRACT
Privacy is one of the important issues now days as privacy is linked with multidimensional issues; security, sentiment, fear, emotions, threats etc. Protecting privacy is as much as data utilization. In this day and age, data is getting generated largely by various industries. Medical industry is one of them. Providing safe access controls and privacy preservation are the primary concerns in the development of medical applications. Medical data possess sensitive information. According to the author, privacy should be preserved at all levels; storage level, to view level to knowledge discovery level. At view level, very limited approaches are proposed to protect the privacy of the medical data. This paper implements Fuzzy C means approach to protect the sensitive data while viewing blood donor data online. In this paper, a sample blood donor records are extracted to categorize the data into high sensitive data and low sensitive data using fuzzy C means rules. Subsequently, the model teaches the underlying relations to perform categorization based on the input. This paper describes the experiment in view of privacy preserving data mining. The experiment is simulated using MATLAB and shows satisfactory result.

Keywords
Sensitive data, Non sensitive data, confidential data, privacy preserving data mining, FCM algorithm.

1. INTRODUCTION
Privacy preservation is vital requirement in hospitals, as they maintain patient's medical records. There are several straightforward controls exist for privacy preservation. Medical industry need to adopt strong tools to protect data and privacy. In addition to the standard controls, hospitals follows privacy and copy right laws to protect individual's privacy. Implementing adequate access (view) controls to hide sensitive data from the third party users is utmost important. There are standard database controls to protect the sensitive data from unauthorized access. Standard database controls are static. They work as per the design. However, one should not protect or hide the data 100%. Consequently utilization of the data hampers. So, machines should be intelligent and dynamic to decide what is to be hidden and what is not to be hidden based on the user input. This paper explains the model using Fuzzy rules. The model hides blood donor data (sensitive data) from third party users while retrieving the blood group information. The proposed system takes input, passes the input to the fuzzy system (rules) to retrieve the output. Fuzzy C means algorithm is one of the well-known algorithms of data mining. The privacy preservation is necessary in data mining applications as it disclose knowledge.

Privacy preservation is important in all most all applications. Recent advances in data collection, data dissemination and related technologies have inaugurated a new era of research called Privacy preserving Data Mining (PPDM). In Privacy Preserving Data Mining (PPDM), the goal is to perform data mining operations on sets of data without disclosing the contents of the sensitive data.

In knowledge discovery, data mining techniques and algorithms play an essential role. So as to construct precise models, data miners frequently need full access to the data [1]. Data mining can potentially help all physicians in a range of ways, through helping interpret complex diagnostic tests, by combining information from multiple sources like sample movies, images, clinical data, proteomics, scientific knowledge, by providing support for differential diagnosis and providing patient-specific prognosis [3]. Data mining is predominantly susceptible to misuse with its promise to resourcefully determine expensive, non-obvious information from large databases [4]. Data mining is performed with large database, where it contains sensitive information [5].

Information mining frameworks that assemble classifiers are one of the routinely utilized devices. An accumulation of cases is taken as contribution by such frameworks, each having a place with one of few classes and portrayed by its qualities for a settled arrangement of characteristics, and yield a classifier that can exactly gauge the class to which another case has a place [2].

For the new information to be consequently gathered and to be added to databases, protection issues are additionally exacerbated by the web [7]. The most vital thought in protection of privacy in information mining is twofold; at first, from the database; delicate basic information like identifiers, names, locations etc are to be modified or trimmed out. Next, since delicate information can similarly well trade off information security, such learning that are mined from a database, by utilizing information mining calculations ought to be avoided [8]. Changing the first information is the central goal of privacy preserving data mining, all together, the private information even after the mining procedure should be non-disclosure able [9][26]

Utility based mining plays a significant role in privacy preserving data mining. To locate the high utility item sets, utility mining is used. User-defined utility is based on the information not available in the transaction dataset. It frequently reflects user preference and can be represented by an exterior utility table or utility function. In a given database, utility table determines the utilities of all items [10]. In a transaction database utility mining discovers all item sets whose utility values are identical to or better than a user specific threshold [11]. Protecting privacy in utility mining minimizes the number of non-sensitive patterns lost [12].

This paper explains the proposed model with an artificial problem scenario followed by solutions.

Problem scenario – Online sensitive blood donor information access. The abstract of this problem is as follows:

Blood bank maintenance is one of the vital and sensitive medical operations. Only hospitals or government agencies can preserve the blood details (blood group, donor details, donor address etc). Often people search for donor information for the needy patients. Hospitals or blood banks provide donor information to the members of consortium hospitals or government agencies based on credentials. Usually to access donor's information, users should provide credentials like password/security code, name of the agency, SSN etc. In the real world, if a person forget password or key, then online systems gives option to re-generate the password and provide the information. However, data mining operations does not focus on regeneration of the password. Without regeneration of the password, the system should be able to retrieve the donors' information, based on the validity of the granular information, intelligently.

The objective of the proposed model is to provide the donors sensitive information like donor name, address, number of times donated, phone number etc., only to the authenticated users, by using fuzzy rules and training.

The proceeding section explains the two approaches for solving the above stated problem scenario; Traditional approach (hard computing) and Fuzzy approach (soft computing)

Traditional approach (Hard Computing)
Traditional approach solves given problem by setting various constraints through hard computing. Hard computing paves penalties in terms of code maintenance, threats, and other software failures. Hard computing works as per the straight forward rules (constraints). At the time of system design, constraints are embedded in the code [25]. These constraints follow binary principle (Yes/No). Decisions based on binary principle should not be the choice always as they maintain only any one of the two possible states. It gives either 100% success or 100% failure.

Fuzzy approach (Soft Computing)
Fuzzy approach adopts commonsense and intelligence approach for solving the classification problems. It follows percentage of truthness; instead of binary (Yes/No) approach. i.e. percentage of membership. In this work, the model retrieves the sensitive information based on the user input by evaluating the truthness based on fuzzy rules and training.

In this work, the donors' information along with blood group is stored in different files t1, t2, t3…..tn. These files are combined and stored in a single large database (called generated database). This data is classified into two types: sensitive data and non-sensitive data by using fuzzy C means algorithm. Further the system is tested with user input and corresponding output is observed. The user input is passed to the trained model. The trained model assigns a score [0.0 to 1.0] to the user input to display the sensitive data. Based on the scores, the model retrieves high sensitive data or low sensitive data. The assignment of score is purely based on underlying model. For example, very low score, 0.2, indicates low sensitive data access. Whereas the score 0.9 indicates high sensitive data access.

The proceeding sections explain the related work, proposed methodology based on Fuzzy Rule and training & testing.

2. RELATED WORKS

Extraction of interesting patterns or knowledge from enormous amount of data is known as data mining. Now-a-days, in data mining data storage and data processing, privacy preservation has been one of the superior concerns with the development of Internet technologies. For privacy preserving data mining, a number of methods and techniques have been developed.

In 2010, Pingshui Wang [13] has provided a broad survey of different privacy preserving data mining algorithms and analyses the representative techniques for privacy preserving data mining, and points out their merits and demerits. In conclusion the author explained problems and instructions for future research.

From a lot of information, information mining manages programmed extraction of already unidentified examples. These informational collections naturally incorporate touchy individual data, which thus get presented to alternate gatherings. They should ensure that information security was kept up in case of information mining in spite of the fact that they can't dismiss the advantages of learning discovery that comes through information mining. Amid information mining protection safeguarding information mining is a committed movement in which the information security is guaranteed. Information protection was as critical as the removed learning and endeavors that guarantee information security in information mining are empowered. In 2010, Mohammad Ali Kadampur et al. [14] have proposed an approach that ensures the information protection amid choice tree examination of information mining process. Behind investigating the choice tree of the first information they have proposed to abut exact commotion to the numeric qualities. For choice tree examination the muddled information was then introduced to the second party. Amid the mining procedure the choice tree got on the first information and the jumbled information are comparative yet by utilizing self composed technique in this work, the information legitimate isn't presented to the second party and thusly the security will be protected. In 2009, Alka Gangrade et al. [16] have addressed privacy-preserving classification problem in a cooperative sense. In a secured manner they have focused the general classification and without the participation of third party Privacy-preserving decision tree classifier using C4.5 algorithm has been introduced. C4.5 algorithm was a software extension of the basic ID3 algorithm designed by Quinlan. Than any existing solutions this protocol was significantly more competent.

In 2010, Jieh-Shan Yeh et al. [23] have proposed Privacy preserving data mining (PPDM). The proposed model is a popular topic in the research community. In the sharing process to hit a balance between privacy protection and knowledge discovery was an important issue. To attain the goal of hiding sensitive item sets so that the adversaries cannot mine them from the modified database; the proposed study focuses on privacy preserving utility mining (PPUM) and presents two novel algorithms, HHUIF and MSICF. On the sanitized database of hiding sensitive item sets the work also minimizes the force. On two synthetic datasets the HHUIF achieves lower miss costs than MSICF is proved by the experimental results. On the other hand, among original and sanitized databases MSICF usually has a lower difference ratio than HHUIF.

In 2011, Mohammad Reza Keyvanpour et al. [15] have presented a Data modification based Framework. This framework is proposed mainly for classification and

assessment of the privacy preserving data mining techniques. The techniques are divided into two major groups based on a framework, explicitly perturbation approach and anonymization approach. Also in proposed framework, to examine and analogically evaluate the techniques in these two major groups eight functional criteria will be used. For more precise comparison of the given techniques the proposed framework provides a good basis to the privacy preserving data mining. Additionally, for different approaches and identifying current approaches in the proposed field the proposed framework allows recognizing the overlapping amount.

In 2011, Vijayarani et al. [22] in their data warehouses, data mining is the extraction of concealed analytical information from large databases and also an influential technology with great potential to examine important information. In the field of data mining privacy preserving data mining was a most recent research area that in general deals with the side effects of the data mining techniques. Protecting individual's information is known as privacy. In data mining research, protection of privacy has become a significant issue. Sensitive outlier protection was novel research in the data mining research field. Division of data into groups of similar objects is said as clustering. Outlier Detection is one of the chief tasks in data mining research. In data mining, for detecting the outliers proficiently clustering algorithms were used. In the proposed technique, to detect outliers they used four clustering algorithms and also proposed a new privacy technique Gaussian Perturbation Random Method to guard the sensitive outliers in health data sets.

In 2011, Gayatri Nayak et al. [19] have discussed method for randomization, k-anonymization, and distributed privacy preserving data mining. The proposed method is less prone to fall prey to the evil hacker sharks of information technology since the Knowledge was supremacy and they are well-informed about information break-in.

In 2011, Archana Tomar et al.[21] have published about the movement in information mining innovation. The innovation is proposed to assess the tremendous measure of information that has assumed an imperative part in a few zones of Business preparing. If not done or utilized appropriately information mining likewise opens new dangers to protection and data security. The main issue was that from non-touchy information, one could conclude delicate data, and individual data, actuality or even examples that were created by any calculation of information mining. The crucial answer for address the issue of security was introduced consecutively to center around protection safeguarding affiliation govern mining. In various research papers the answer for review diverse perspectives were talked about. In the wake of dissecting those examination papers another arrangement was finished up which was best in viability and execution. Before examining the calculations, to build up the more efficacious model the information structure of database and delicate affiliation run mining set have been broke down.

In 2012, Thavavel et al.[17] have proposed a solution to this problem by managing unstructured data into structured data by means of legacy system and distributed data partitioned method that gives distributed data for mining multi text documents. The proposed frame work gives the testing of the similarities among text documents and privacy preserving Meta data hiding technique, which are explored in text mining.

In 2012, W. T. Chembian et al. [18] have managed vast amount of personal data and distribution of these data was proved to be helpful for data mining application. In privacy and security research, privacy-preserving data mining (PPDM) was one of the latest trends. In several cases, unless the privacy of sensitive information is assured, users are reluctant to provide personal information. Because of the huge quantity of consumer data tracked by automated systems on the internet, privacy preserving data mining has turn into an imperative problem in recent years. To provide improved privacy than K-Anonymity method, they have intended a blocking algorithm. In the Blocking based algorithms the design was to alternate the value of an item supporting the rule they desire to hide with a worthless symbol.

In 2012, M. Sridhar et al addressed the problem of data anonymization in data mining, cryptography, and information hiding. The authors described the need of fuzzy attributes to safe guard the sensitive data. The authors addressed privacy preserving data mining using fuzzy logic. The experiment transforms a variable to fuzzy attributes. Fuzzy attributes does not reveal the original data. Hence, individual privacy can be protected in better way.

In 2014, Ravi Sankal et al, proposed a technique to diagnose diabetic patients using Fuzzy C means and SVM approach w.r.t to data mining. The proposed technique is tested using UCI diabetic data set. The proposed method predicts whether a patient possesses diabetic or not. The accuracy level of FCM is 94.3% and positive predicted value is 88.57% [27].

In 2015, G.Rasita banu et al developed predictive system using FCM clustering algorithm for finding the risk of heart attack of a patient using the profiles collected from the patients. The model effectively classified abnormal and normal cases. Initially pre -processing of the data is done to remove all the duplicate records and add missing data. In the classification stage, a FCM classifier is used to classify the data as heart disease present or not. The results of classification experiment, preformed over data sets obtained from 270 patients, shows that the classifier has achieved better accuracy. The performance of the proposed FCM is proved to be a well known approach in terms of accuracy [28]

In 2018, Ruby Bhuvan Jain et al, proposed a dynamic data masking model to protect sensitive data using non-deterministic random replacement algorithm. Masking is one of the techniques for data replacement. The authors proposed a non-deterministic random replacement method. Using this method the masked data cannot be reversed back to original data. The authors conducted experiment on bank data. This method primaly useful to protect the data from attacks. The proposed method is a preventive technique. This method applicable as a pre-task for data sensitivity preservation and applicable for structured data only [25]

3. METHODOLOGY

This paper describes design & implementation to classify the blood donor database into two types of data sets; Sensitive dataset and low-sensitive dataset. Low sensitive dataset could be viewed by any individual. Whereas, the high sensitive datasets can be viewed by authorized person only. For example, patient diagnosis information is high sensitive information. The figure below shows the abstract view of the working process of privacy preserving of medical data related to the blood bank information.

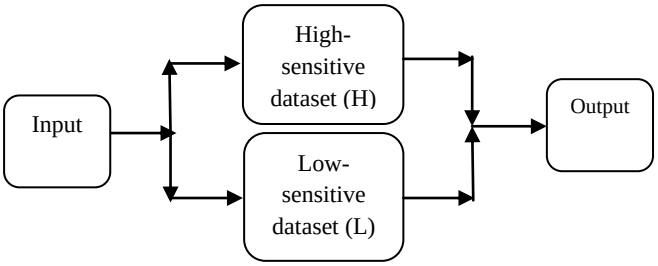


Fig 1. Abstract view - Fuzzy rule classification

At the beginning, the process generates the blood bank data set, for which, existing donor information files are combined together to form a dataset which is stored in main database and is represented as follows :

$$T = t_1, t_2, t_3 \ldots\ldots t_n$$

T = Total files in the main database

$t_1, t_2, t_3 \ldots\ldots t_n$ = Individual files in the main dataset

The main process of this work is clearly explained below in three steps:

(i) Feature Extraction

(ii) Fuzzy Rules Generation

(iii) Network Training and Testing

(i) **Feature Extraction**

The features extraction is a method to extract feature data from the main database for training process for the betterment of output to be obtained. The following constraints are assumed to generate rule using Fuzzy C means algorithm. Threshold values are set for each combinational value. See the following:

Constraint (1): If Bg = 1 and P = 0, for example the value assigned for this process is 0.3. Actual Threshold value is 0.5. The obtained value 0.3 < Threshold value, so that the feature extraction is mentioned as "**L**". Similarly, under the similar conditions, for the remaining records, the extraction process assigns a value less than 0.5 and greater than 0.0.

Constraint (2): If Bg = 1 and P = 1, for example the value assigned for this process is 0.7. Actual Threshold value is 0.5.The obtained value 0.7 > Threshold value, so that the feature extraction is mentioned as "H". Similarly, under the similar conditions, for the remaining records, the extraction process assigns a value greater than 0.5 and less than 1.

In the above constraints, Bg = 1; indicate that Bg for Blood group and 1 indicate that it contains some value."P" stands for password, if it is "0" then it indicates password is blank and if it is "1" then it contains a value.

(ii) Fuzzy Rule Generation Using C-Means Algorithm

A fuzzy rule is a series of if-then else rules. Fuzzy Logic was introduced in 1965 by Lofti A. He is considered as the father of Fuzzy Logic. Fuzzy logic is logic system. In which membership is a matter of degree. In this paper, membership indicates, the level of sensitive data. You can use Fuzzy Logic Toolbox software with MATLAB® technical computing software as a tool for solving problems with fuzzy logic. Fuzzy logic is a fascinating area of research because it does a good job of trading off between significance and precision—something that humans have been managing for a very long time. Lotfi Zadeh, who is considered to be the father of fuzzy logic, once remarked: "In almost every case you can build the same product without fuzzy logic, but fuzzy is faster and cheaper." The basic working of this algorithm is as follows:

"Fuzzy C means works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center, more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration, membership and cluster centers are updated according to the formula" [24]. The number of clusters assumed in this work are 3.

Steps by step procedure of Fuzzy C-means Algorithm

Let, $T = t_1, t_2, t_3 \ldots\ldots t_n$

$C = C_1, C_2, t_C \ldots\ldots C_c$ is the set of centers

Step 1

Select 'c' cluster center randomly

Step 2

In this compute fuzzy membership ψ_{xy} using below formula

$$\psi_{xy} = \frac{1}{\sum_{m=1}^{c} \left(d_{xy} / d_{xm} \right)^{2/K-1}}$$

Step 3

Compute Fuzzy center ς_y using below formula

$$\varsigma_y = \frac{\left(\sum_{x=1}^{n} (\psi_{xy})^k \, z_x \right)}{\left(\sum_{x=1}^{n} (\psi_{xy})^k \right)} \forall_y = 1,2,3,.....c$$

Step 4

Repeat (3) & (4) till minimum value for 'E' is obtained, is also said to be $\left\| H^{(m+1)} - H^{(m)} \right\| < \alpha$

Where,

"M" is the iteration step,

"α" is the termination criterion between [0, 1],

$H = (\psi_{xy})_{n*c}$ Is the Fuzzy membership matrix

"E" is the objective function.

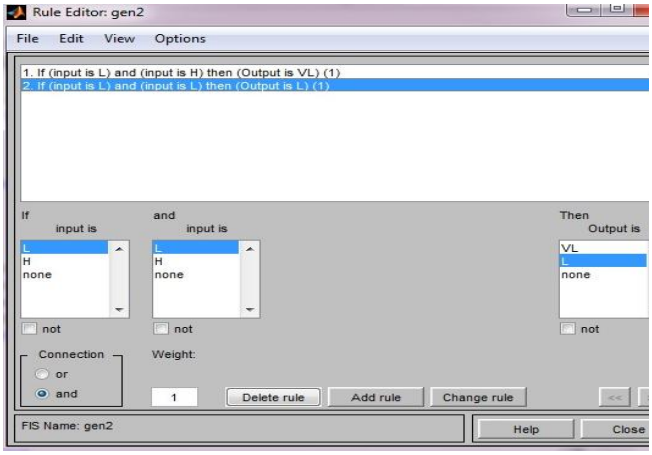


Fig 2. (Fuzzy rule generation)

The aforementioned fig 2 shows the rule generation. It shows that for first condition, if input L and H are chosen the output will be VL (very low). In the second rule L (low) and L (low) are chosen, and then the output is L. If both inputs are H (high) and H (high) then the resultant would be high.

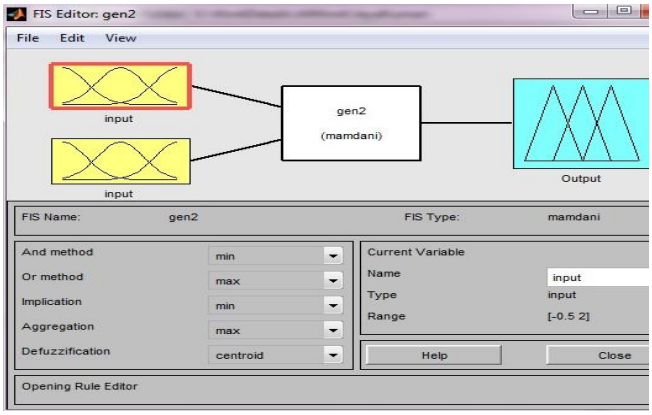
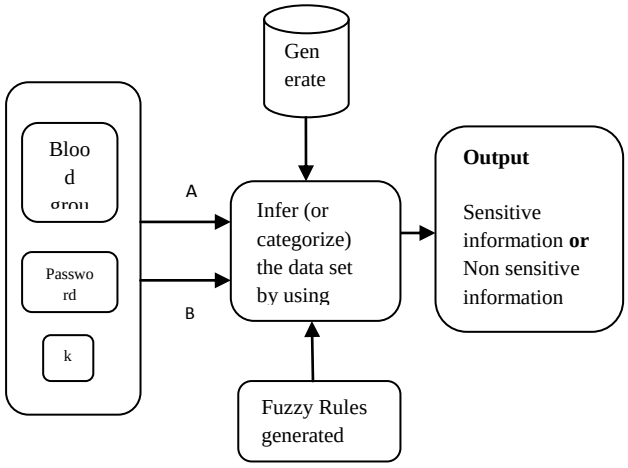


Fig 3. (Fuzzy output)

The above figure 3 shows graph , which indicates the input wave form (left side) . Whereas the graph on the right side indicates the output wave form. Wave form representation is the standard method of representing input and output.



iii) Network Training and Testing

the above fig 4 shows that the generated database contains both sensitive and non-sensitive classified data. The fuzzy system is tested with new user input. Here, a user may be authorized user or third party user: Authorized user possess password and third party user does not possess password. The fig 4 represents input, classification and output areas. Input area contains many fields like username, organization name, hospital name; SSN, consortium_ID etc. Due to space limitation we have shown only blood group, password and Ok buttons in the input area.

The training process is based on the following two cases:

Case 1

If blood group name is given in the blood group section shown in the aforementioned fig.1 without password. The given blood group name send via "A" transmitting path for the next level processing. If the input given without password, the transmitting channel assign a value less than threshold value which pave path to access to low-sensitive database.

Case 2

Along with the blood group name, if password entered, this paves the transmission towards high-sensitive data. i.e it transmit the input data through path B.

4. RESULTS AND DISCUSSION

The following section explains the input and output with detailed explanation**.**

Input form without password

The below shown fig 5 represents a part of the input; blood group name and without password.

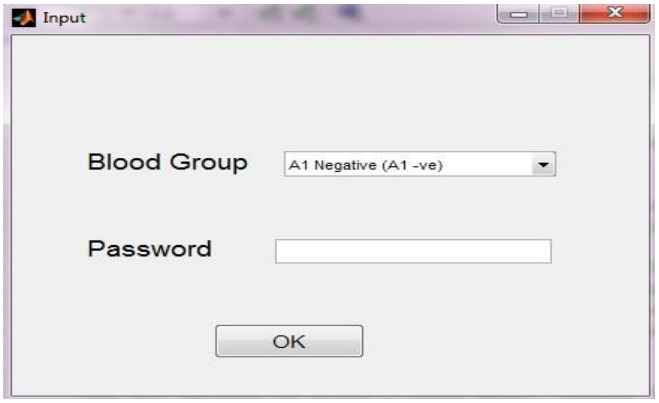


Fig 5. (Initial form)

Regarding to the input determined is clearly shown in above figure i.e. Blood group name and enter OK button, the retrieved display is shown in below figure 6,having blood group name and total count that particular blood group have. This indicate that it preserve the sensitive data and shows only the non-sensitive data to the third party users. The sensitive information such as name of the donators, address and the number of count that particular person donate, these are all the information which are hide to the third party users. Output retrieved for the correspondent input without password

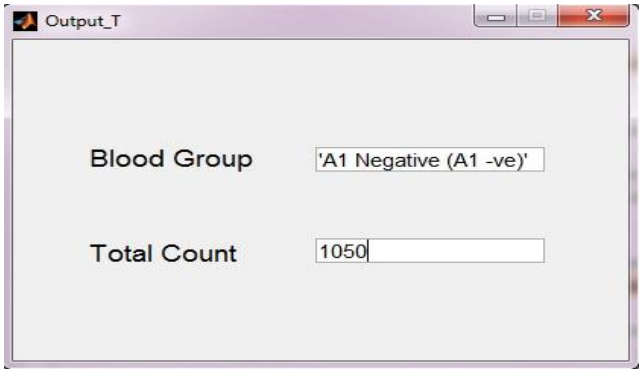


Fig 6. (Output retrieval for the given input form)

The aforementioned figure 6 indicate blood group name and total blood count in the blood bank which is not more sensitive, when compare to other information in the blood bank dataset. Input form with password

Fig 7 (Second step for the initial process)

In this password is given along with the blood group name, this paves the path towards main dataset and it retrieve the information of all which are present in the main dataset. Output retrieved for the corresponding input with password

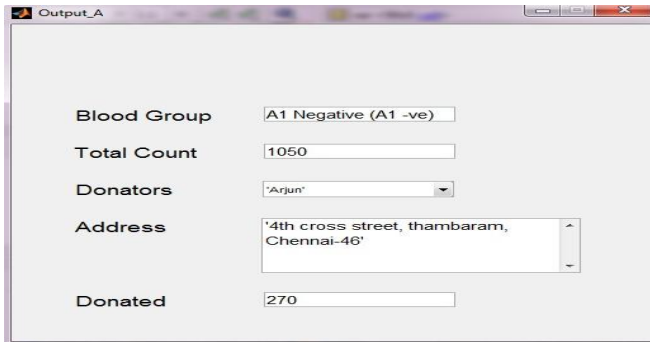


Fig 8. (Output retrieved for the second step for the initial process input)

In the above mentioned Figure 8 which retrieve the information such as blood group name, total count, donator name, donar Address, how much the particular person donates. These are all the information retrieved, if passwd entered. Donator name, Donator Address, and unit the donator donates are considered to be sensitive data. This sort of information is retrieved based on the fuzzy rules based on the training database (Here called generated database). The fig 8 represents list of donors under list box. If we click on the list box, then it shows the names of the donors having similar blood group. There can be more than one user having similar blood group. Whichever the donor name clicked, the corresponding address and number of times donated are displayed.

5. CONCLUSIONS

In this work an efficient technique fuzzy c-means is utilized to generate rules to classify sensitive data and non- sensitive data. Fuzzy c means is one of the frequently used data mining tasks. Here, the model is designed in such a way that based on input feed, output will be retrieved. In this work, Password is one of the constraints to retrieve the sensitive and non-sensitive retrieval. In reality, a user may forget the password and other credentials; even then the system should be able to retrieve the sensitive information based on the amount of input correctly fed. In future, the author wishes to implement the same experiment by using Neural Network (NN) techniques and compare the results with Fuzzy C means for performance analysis.

6. ACKNOWLEDGMENTS

Sincere thanks to Abacus Institute of Computer Applications, Research Center, SPPU, Pune for giving opportunity to pursue research related to data privacy.

7. REFERENCES

[1] Mohammad Saad Al-Ahmadi, "Privacy-Preserving Data Mining for Horizontally-Distributed Datasets using EGADP", Journal of communications of the IBIMA, Vol. 5, No.2, pp. 7-15, 2008

[2] Xindong Wu ,Vipin Kumar ,J. Ross Quinlan ,Joydeep Ghosh ,Qiang Yang ,Hiroshi Motoda , Geoffrey J. McLachlan , Angus Ng , Bing Liu , Philip S. Yu Zhi-Hua Zhou , Michael Steinbach , David J. Hand Dan Steinberg, "Top 10 algorithms in data mining",3,Vol.14,pp.1-37,2007

[3] Florin Gorunescu,"Data Mining Techniques in Computer-Aided Diagnosis: Non-Invasive Cancer Detection", World Academy of Science, Engineering and Technology, Vol.34, pp.280-283, 2007

[4] Alexandre Evfimievski, Johannes Gehrke and Ramakrishna Srikant, "Limiting Privacy Breaches in Privacy Preserving Data Mining", In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database system, pp. 211-222, 2003

[5] Deepika Saxena, "Privacy of Data, Preserving in Data Mining", International Journal of Scientific & Engineering Research, Vol. 2, No. 3, pp. 1-5, March 2011

[6] Srinivasa Rao and Chiranjeevi, "Distortion Based Algorithms for Privacy Preserving Frequent Item Set Mining ", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.1, No.4, pp. 16-27, July 2011

[7] Evfimievski, Srikant, Agrawal and Gehrke, "Privacy preserving mining of association rules". In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Canada, pp. 1-12, 2002

[8] Isaac Cano, Susana Ladra, Vicenc and Torra, "Evaluation of Information Loss for Privacy Preserving Data Mining through comparison of Fuzzy Partitions", In Proceedings of the IEEE World Conference on Computational Intelligence, Barcelona (Espana), 2010

[9] Vassilios S. Verykios, Elisa Bertino, Igor Nai Fovino, Loredana Parasiliti Provenza, Yucel Saygin and annis Theodoridis, "State-of-the-art in Privacy Preserving Data Mining", Newsletter ACM SIGMOD Record, Vol. 33, no.1, pp. 1-8, 2004

[10] Vid Podpecan, Nada Lavrac and Igor Kononenko, "A Fast Algorithm for Mining Utility-Frequent Itemsets", In Proceedings of the Eleventh European Conference on Principles and Practice of Knowledge Discovery in Databases, 2007

[11] Ciriani, De Capitani di Vimercati, Foresti, and Samarati, "Chapter 1 K-Anonymous Data Mining: A Survey", Springer, pp. 1-34, 2008

[12] Rajalaxmi and Nataraja, "An Effective Data Sanitization Approach for Privacy Preserving Utility Itemset Mining", International Journal of Engineering Research & Industrial. Applications, Vol.1, No. 6, pp 133-143, 2008

[13] Pingshui WANG, "Survey on Privacy Preserving Data Mining", International Journal of Digital Content Technology and its Applications, Vol.4, No.9, pp.1-7, 2010

[14] Mohammad Ali Kadampur and Somayajulu, "A Noise Addition Scheme in Decision Tree for Privacy Preserving Data Mining", Journal of Computing, Vol.2, No.1, pp.137-144, 2010

[15] Mohammad Reza Keyvanpour and Somayyeh Seifi Moradi, "Classification and Evaluation the Privacy Preserving Data Mining Techniques by using a Data Modification–based Framework", International Journal on Computer Science and Engineering, Vol.3, No.2, pp.862-870, 2011

[16] Alka Gangrade and Ravindra Patel, "Building Privacy-Preserving C4.5 Decision Tree Classifier On Multi-Parties" ,International Journal on Computer Science and Engineering,Vol.1,No.1,pp.199-205,2009

[17] V.Thavavel and S.Sivakumar, "A generalized Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment ",International Journal of Computer Science Issues,Vol.9,No.2,pp.434-441,2012

[18] W. T. Chembia and,J. Janet ,"An Efficient Randomization Algorithm for Privacy Preserving Data Mining" ,European Journal of Scientific Research ,Vol.77,No.3,pp.303-308,2012

[19] Gayatri Nayak and Swagatika Devi, "A Survey on Privacy Preserving Data Mining: Approaches and Techniques", International Journal of Engineering Science and Technology, Vol.3, No.3, pp.2127-2133, 2011

[20] Balamurugan,J. Bhuvana and S. Chenthur Pandian, "Privacy Preserved Collaborative Secure Multiparty Data Mining ",Journal of Computer Science,Vol.6,No.6,pp.872-878,2012

[21] Archana Tomar,Vineet Richhariya and R.K. Pandey,"A Comprehensive Survey of Privacy Preserving Algorithm of Association Rule Mining in Centralized Database" ,International Journal of Computer Applications,Vol.16,No.5,pp.23-27,2011

[22] S.Vijayarani and S.Nithya, "Sensitive Outlier Protection in Privacy Preserving Data Mining", International Journal of Computer Applications, Vol.33, No.3, pp.19-27, 2011

[23] Jieh-Shan Yeh and Po-Chiang Hsu, "HHUIF and MSICF: Novel algorithms for privacy preserving utility mining", Expert Systems with Applications, Vol.37, pp.4779-4786, 2010

[24]Https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm

[25] Ruby Bhuvan Jain, Manimala Puri and Umesh Jain, "A Robust Approach to Secure Structured Sensitive Data using Non-Deterministic Random Replacement Algorithm", International Journal of Computer Applications 179(50):17-21, June 2018.

[26] Ruby Bhuvan Jain, Dr. Manimala Puri, Umesh Jain, "A Robust Dynamic Data Masking Transformation approach To Safeguard Sensitive Data", International Journal on Future Revolution in Computer Science & Communication Engineering, Volume: 4 Issue: 2 ISSN: 2454-4248.Pg no. 366-370.

[26] M. Sridhar et al, Dr.Ravindra Babu, "A Fuzzy Approach for Privacy Preserving in Data Mining", International Journal of Computer Applications (IJCA), 2012, Vol57.no.18.

[27] Ravi Sankar et al, "Prognosis of Diabetes Using Data mining Approach- Fuzzy C Means Clustering and Support Vector Machine", International Journal of Computer Trends and Technology", Vol.11, no.2, May 2014.

[28] Rashita Banu et al, " Predicting Heart attack using Fuzzy C means clustering algorithm", International Journal of Latest trends in Engineering and Technology, 2015, Vol.5, No.3.