Prevention of Sensitive and Privileged Information Using Fuzzy C-Means with DES Approach

Kishore Mishra
Computer Science & Engineering
Apex Institute of Engineering And Technology
Jaipur, India

Associate Prof. Brij Kishore
Computer Science & Engineering
Apex Institute of Engineering And Technology
Jaipur, India

***Abstract*— This time privacy preserving is the most concerning issue. Anonymity continues the identification of the individuals in competencies methods personal; however it's not always concerned with how public the traces hence come to be. Anonymization method is applied on the dataset. Then by using of fuzzy c-means clustering data are clustered. Finally, on this data apply DES technique to preserve privacy. An experimental result shows that fuzzy c-means with DES encryption algorithm gives most accurate results with reduced error rate rather than previous approach.**

Keywords*— *Data Mining, Fuzzy C-Means Clustering, Privcy preserving, Data encryption standard technique, Anonymity.

I. INTRODUCTION

Data mining is, *the removal of hidden predictive information from extensive database*, is a powerful new innovation with incredible potential to enable organizations centered around the most vital info in their data warehouses [1].

Privacy preserving data mining (PPDM) is partitioned into differs classification. We will survey the essential idea of PPDM and diverse examination performed in the area of PPDM under different classification. We will center around estimations that are used to measure the responses happened in light of privacy preserving technique [2].

Data Anonymization likewise alluded as data jumbling, data veiling, de-sharpening, de-distinguishing proof or data scouring) is the procedure that covers private information. It secures sensitive information underway data base so it can be exchanged to a test group. Data anonymization can be ordered to unadulterated anonymization and pseudo-anonymization [3]. To give a safe dataset, the publisher can change the information such that they can ensure that no assailant can induce any individual data of the enclosed individual records. Given privacy guarantees offer different privacy levels and deal with different attack models. The procedure of anonymization in every one of these cases requires some fundamental information alterations to be performed, with a specific end goal to guarantee the fulfillment of the given security guarantee [4].

II. LITERATURE SURVEY

K. Sashirekha et al (2014) displayed that, Privacy Preserving and the Data Mining tends to the issues of securing mobile entities from the invaders. Privacy threat includes process of predicting the pattern movement based on statistical information collected. Intruder displays the models of traffic to predict cluster movement and try to access the private info of mobile users. Privacy can be achieved by the means of randomization, k-anonymization and distributed privacy-preserving data mining. To give better privacy multi-level frameworks are used. Here, an examination done on various technique of the privacy preserving and policy of multi-level trust, limitation while utilizing enormous dimension data sets [5].

Prakash et al. (2015) a customized anonymization technique is recommended which monitors the security while the private data is available. The major contributions of this paper are three folds: (i) the definition of the data gathering and publication method, (ii) the privacy framework model and (iii) personalized anonymization methodology. The investigational study is obtainable at the last part; it demonstrates this method executes enhanced over the distinct l-diversity measure, probabilistic l-diversity measure and k-anonymity with t-closeness measure [6].

Yanmin Gong et al. (2015) recommend a protection saving framework for IDR programs in the smart grid, which empowers the DR provider to process singular request curtailments and DR rewards while protecting client privacy. Moreover, a customer can reveal his/her identity and prove ownership of his/her power usage pro- file in certain situations, such as legal disputes. We accomplish both security and productivity in our plan through a blend of a few cryptographic natives, for example, personality committable and signatures and partially blind signatures. As far as we know, we are the first to identify and address privacy issues for IDR programs in the smart grid [7].

Samet Tonyali et al. (2015) suggest to execute a meter data obfuscation method to preserve customer privacy that has the capability to perform dissemination state estimation. We at that point assess its execution on a broad Advanced Metering Infrastructure (AMI) organize based upon the new IEEE 802.11s remote work standard. For the data obfuscation method, we suggest two protected obfuscation value distribution mechanisms on this 802.11s-based remote network. Utilizing confusion esteems gave through this strategy, the meter readings are jumbled to safeguard costumer security from eavesdroppers and the adequacy organizations while preserving the utility companies' capability to use the data for state valuation. We measured the impact of this method on data good put, delay and packet delivery ratio under a variety of situations. Simulation outcomes have presented that the suggested method can offer

very alike performance to that of non-privacy method with insignificant expenses on the meters and network [8].

Gursoy et al. (2016) objectives to gauge such strategies on learning investigation by moving toward the issue from two recognitions: (1) the information is anonymized and afterward imparted to a learning examination master, and (2) the learning examination dexterous is given a protection saving interface that oversees her entrance to the information. We develop proof-of-concept executions of privacy preserving learning analytics jobs using both perceptions and run them on syntheticand datasets. We also represent an experimental study on the trade-off between entities' privacy and the accurateness of the learning analytics jobs [9].

Putri et al. (2016) prescribes a crossover change in PPDM, which is a combination of the two existing procedures on prior investigations, the entropy-based segment method and collective distortion techniques. To measure the recommended method, estimation of the utility and confidential parameter estimation are used. Utility evaluation is used to assess the accuracy of the information and privacy parameter evaluation to assess how close the original value will be obtained from the transformation and how much they are distorted. The experimental outcomes show that the recommended method provides better outcomes than earlier methods in utility and confidentiality, so the information shall be preserved and can be used for analyzing such as data mining [10].

Peter Shaojui Wang et al. (2016) consider another insider hazard for the security safeguarding work of disseminated kernel-based data mining (DKBDM), for example distributed SVM. Once deliberated a negligible concern, insider attacks have grown to be one of the main three vital data violations. A flawed structure empowers agreement to go unnoticed when an insider profits information with an outcast, who would then be able to enhance the special data from message transmissions (delegate part esteems) among foundation. To the best of our insight, we are the first to find this new insider risk in DKBDM. We also analytically demonstrate the minimum amount of insider data necessary to launch the insider attack. At long last, we follow up by presenting a few proposed privacy-preserving plans to counter the portrayed assault [11].

III. Proposed Methodology

A. Fuzzy C-Means Clustering

The FCM operates fuzzy divisioning to such a degree, to the point that an data point can have a place with whole gatherings with various membership range between 0 & 1. This algo works by consigning membership to all data point involving towards each group center based on distance between the cluster center & the data point. Increasingly, the data is close to the group focus more is its membership towards the particular cluster center. Clearly, summation of participation of every data point ought to be equal to one. After, every cycle iteration membership & cluster centers are updated based on formula [12].

B. Data Encryption Standard Approach

The DES is a symmetric-key block cipher easily reached by the NIST. DES is an execution of a Feistel Cipher. It employs 16 round Feistel structure. The block size is 64-bit. However, key length is 64-bit, DES has an proficient key size of 56 bits, since 8 of the 64 bits of the key are not utilized via the encryption algo (work as check bits only). At the encryption site, DES takes a 64-bit plaintext and makes a 64-bit ciphertext; at the disentangling site, DES takes a 64-bit ciphertext and makes a 64-bit block of plaintext. The similar 56-bit cipher key is used for both encryption and decryption.

The proposed algorithm is a try to present a new method for complex encrypting and decrypting data based totally on parallel programming in the sort of way that the new method can make use of multiple- core processor to acquire higher speed with better degree of protection.

Proposed Algorithm:

1. Partition of given dataset through the usage of Fuzzy C- means clustering
2. Summary ← partition
3. Dim ← choose dimension()
4. P ← frequency set(partition , dim)
5. Sv ← find median(P)
6. Lhs ← {t belongs to partition : t.dim <= sv}
   Rhs ← {t belongs to partition: t.dim > sv}
7. Apply DES for encrypting the records.
8. Finally return union of lhs and rhs partition
9. Stop.

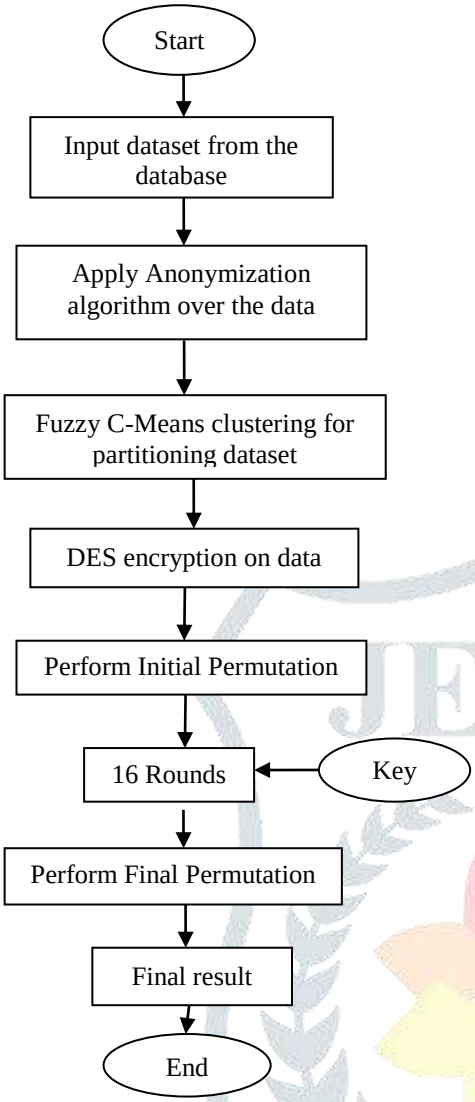

Fig. 3.1 Workflow of Proposed Work

IV. Result Analysis

In the result analysis, the experiment of proposed work performed by using MATLAB tool. This firstly loaded dataset then choose dataset100. Then, gives encrypted form of dataset. Finally get decrypted information through anonymization algorithm then by applying DES encryption algorithm get most accurate results with reduced error rate.

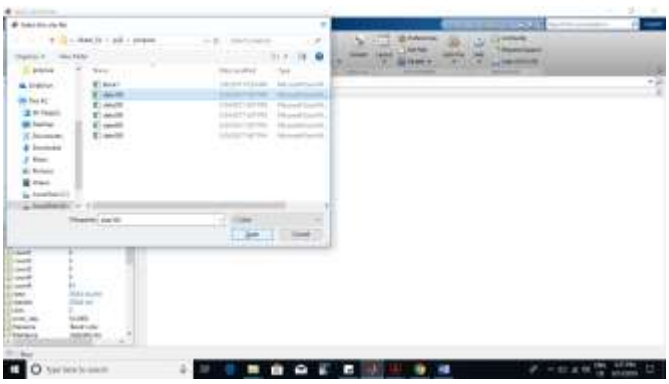

Fig. 4.1 Load dataset

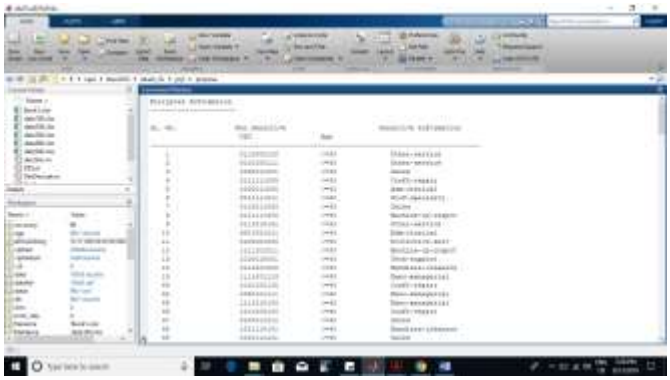

Fig. 4.2 Encrypted Information

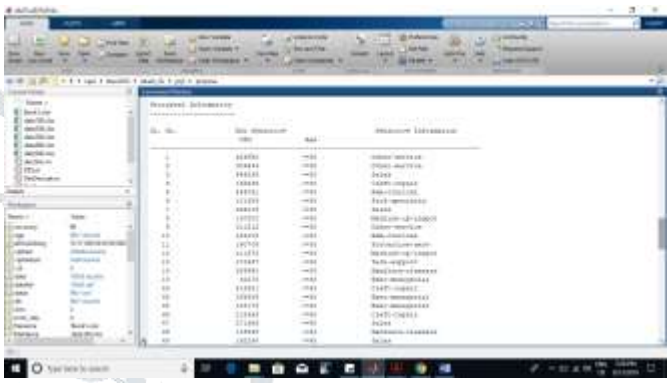

Fig. 4.3 Decrypted Information

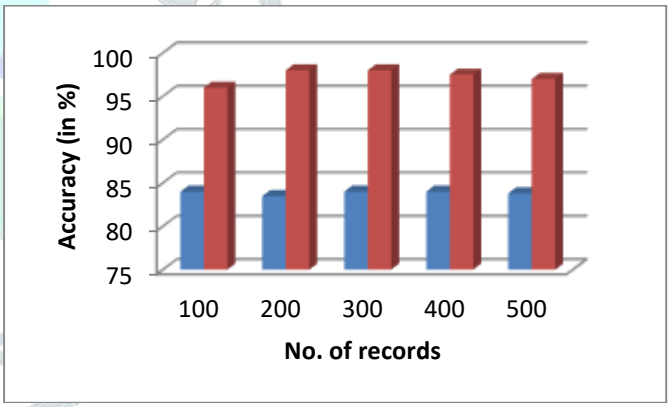

Fig. 4.4 Comparison graph of accuracy in base and propose method results

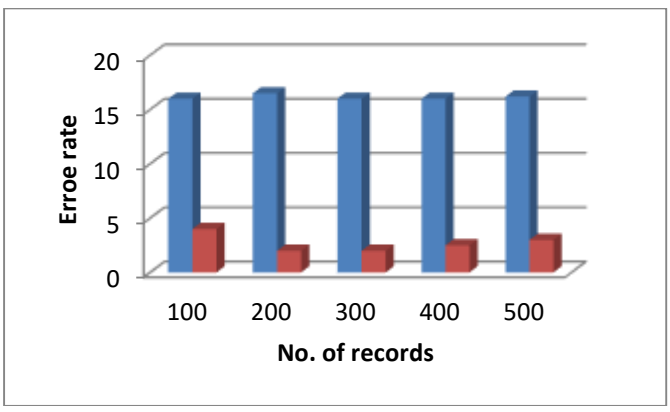

Fig. 4.5 The analysis of error rate in base and propose method results

V. CONCLUSION

Every privacy preserving strategy has its own particular significance. Anonymizing huge data and dealing with anonymized data sets are nonetheless challenges for classic anonymization processes. To preserve privacy, model of k-anonymity has been proposed by the Sweeney which achieve k anonymity by means of generalization and the Suppression, K-anonymity, it is troublesome for a faker to choose the character of the people in accumulation of data set containing individual information.Fuzzy C-mean is an unsupervised, non-deterministic, numerical, iterative methodology for social occasion. This calculation works by doling out membership to every data point relating toward each group distance based on remove between the cluster center and the data point.

The results of our proposed work shows that by doing fuzzy c-means clustering and encrypting the data using DES method we can achieve more preservation of privacy.

References

[1] Mohammadian, M., "Intelligent Agents for Data Mining and Information Retrieval," Hershey, PA Idea Group Publishing, 2004.

[2] Dhivakar K , Mohana S "A Survey on Privacy Preservation Recent Approaches and Techniques" International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2014.

[3] Sergey Vinogradov , Alexander Pastsyak , "Evaluation of Data Anonymization Tools" , IARIA, 2012.

[4] Olga Gkountouna, "A Survey on Privacy Preservation Methods" , June, 2011.

[5] K.Sashirekha, B.A.Sabarish, Arockia Selvaraj, "A Study on Privacy Preserving Data Mining" ISSN(Online):2320-9801July 2014

[6] M. Prakash, G. Singaravel, "An approach for prevention of privacy breach and information leakage in sensitive data mining", Computers and Electrical Engineering xxx (2015).

[7] Yanmin Gong, Ying Cai, Yuanxiong Guo, and Yuguang Fang"A Privacy-Preserving Scheme for Incentive-Based Demand Response in the Smart Grid" IEEE Transactions On Smart Grid, 2015.

[8] Samet Tonyali, Ozan Cakmak, Kemal Akkaya, Mohamed Mahmoud and Ismail Guvenc "Secure Data Obfuscation Scheme to Enable Privacy-Preserving State Estimation in Smart Grid AMI Networks", IEEE, 2015.

[9] Mehmet Emre Gursoy, Ali Inan, Mehmet Ercan Nergiz and Yucel Saygin "Privacy-Preserving Learning Analytics: Challenges and Techniques" IEEE Transactions On Learning Technologies, 2016.

[10] Putri, A walia W., Laksmiwati Hira "Hybrid Transformation in Privacy-Preserving Data Mining", IEEE, 2016.

[11] Peter Shaojui Wang, Feipei Lai, Hsu-Chun Hsiao, And Ja-Ling Wu "Insider Collusion Attack on Privacy-Preserving Kernel-Based Data Mining Systems", 2016.

[12] R.Suganya and R.Shanthi, "Fuzzy C- Means Algorithm- A Review", International Journal of Scientific and Research Publications, Volume 2, Issue 11, November 2012, pp. 1-3.