Hierarchical Privacy Preserving Of High Dimensional Data Using Modified K-Means & Vq Approach

[1]Ms. N. REVATHI, [2]Mrs. INDUJA

[1]Research Scholar, [2]Assistant Professor, PG & Research Department of Computer science, Tirupur Kumaran College for Women, Tirupur, India.

Abstract: - Data mining agreements with a large database of important information. Data preparation requires disclosure of confidential information and forms of confidentiality and privacy obligations. Improve efficient data mining has increased the risk of exposure to key data. When an intimidation data is collected for an individual's privacy, anyone can identify specific people to access the data subway or the newly compiled data structure, especially when data is anonymous. Providing security to key data against an unrecognized approach is a long-term goal for the database security research community and government statistical agencies. Clustering is the protection of privacy by protecting the basic attribute values of clustering analysis. Individuals' privacy is protected by doing so.

This article presents a practical data release framework for creating a distinct version that protects a separate privacy and information analysis for the Guardian's analysis. The quality of the cluster is much better than the quality of the data being hidden without such attention, by focusing on protecting the cluster system in the indirect process of experiments in real-life standards.

Keywords: - Privacy Preserving, Clustering, VQ Model, Security System.

I. INTRODUCTION

Privacy preserving data processing is a new research on data mining and actual databases, destroying the data tunnels by their data privacy. The primary idea of protecting the privacy of data tunnel is twice that. First, important crude data that needs to be altered or cleaned from the primary database of identifiers, names, addresses, etc. must be linked together with the beneficiaries of the data without the privacy of others. Second, important information that can be derived from a database by using data mining techniques must be rejected as well as a similar learning, as well as proving data confidentiality. The basic purpose of privacy to protect the data mining is the process of changing the primary data, with the intention of personal data and personal learning to remain private even after the mining. The problem is when it comes to receiving confidential information from discharged data by unidentified customers, usually called the "turning the database" problem. In this report, we provide a classification and detailed description of the different technologies and systems manufactured in the privacy area that protect data mining.

II. LITERATURE REVIEW

Data mining is widely used to find knowledge from large databases. The problem of data mining can withstand important information that can be disclosed with the receipt of non-core information [5]. As a result, privacy is increasingly important in many data mining applications. The first approach protects the privacy of data by using an access control approach that can be used to secure a privacy policy. Another approach uses encryption methods. [7]. E is used in many areas such as health, cloud, location services, and wireless sensor networks. [10].

K. Venkatesh and Md. Ahamed (2017), it is proposed to protect large data mining and security data processing procedures and techniques. A plan for data retention and greater processing and computational efficiency may be supported to regenerate databases. Data surveillance program may be considered to supervise databases. Data retention of data attribute can be restored to databases from its predecessor databases. It goes away from retrieving databases. In the proposed model, the use of security portrayal application encourages simplified picture of different security issues and a database will be anonymous to different unknown data.

Antorweep Chakravorty, Tomasz Wlodarczyk, and Chunming Rong (2013), proposed, full data lifetime: One way to accomplish data protection and security through data age / collection, exchange, storage, handling, and sharing. The idea of protection from nations, communities and borders. Anyway, everything is related to collection, storage, use, handling, sharing or deletion and securing

 DOI : 10.35291/2454-9150.2019.0226

data from identifiable data. Sen and others. Total data for cloud computing reviews data security and security issues around life cycle. Considering their structure, we determine four areas to ensure security and safety for the intelligent home exposure arrangement.

Sarangkumar S. Dubey, Prof. Seema B. Rathod (2016), Proposed strategy for privacy data for large data processing, or, in other words, and useful strategy for mining information from extensive and corrupted databases using the Hadobo architecture. The HDFS is one of the core architectures that mainly helps us to handle our large data. Hadoop provides a formal record system and a framework for analyzing and modifying the wide range of data sets using the Macpratus world view. An important nature of Hatoyo is to pass data and calculations to thousands of patrons, and execute application calculations equivalent to their rankings. A Hadoop Cluster Scales calculation range, storage range and I / O frequency include item servers.

Anurag, Deepak Arora and Upendra Kumar (2018) proposed, Data mining systems that protect privacy rights between UML graphic model model run time requirements to improve global designers. For example, the distribution center that combines other information with the database, login certificates. Data Mining Server's functionality is an object oriented program in the leading and query language in the backend, and is expected to be introduced by both Object Oriented Programming and Question Package status on the data center server. The questions are sent to the data warehouse server. The username and password required for the login credentials database is required and the connection to the useful verification has occurred. Data comes with data from stockroom database and is attached to privacy security tasks.

A. N. K. Zaman, and Charlie Obimbo (2014) proposed, For example, continuous data is the most useful for a wide velocity filter, ranging from the classical vector, to a highly configurable structure for the highest ranking. Data mining (PPDM) and PPDP, which protect privacy, are found in literature in any event. Data-based decision making processes have been confirmed by data mining in expanding data repositories in the past with corporations and governments. Scientific classification tree shows two attributes age and disease code.

III. PROPOSED METHODOLOGY

There are distinct types of sample in an information package. These models may be a mixture of information mining processes, for example, arrangement, collection, connection base mining, exception testing, and development analysis. On many occasions, information mining efforts have to be made in the information published by the customer to find a consumer model. Ideally, a security stabilization system should keep the information quality to help mining and measurable experiment. Nonetheless, it only collects data mining for information collection. Security sources may focus around the tasks that support them.

Information Technology is described as a means of discovering real, new connecting pioneers and patterns by reviewing them every time, using cloud specification, provision of sampling using the model's authentication, and number procedures. If a built-in query language (eg) has access to most information or a specific information, information collectors will tell you exactly what they need, even if they do not use any information. Consequently, a SQL query is generally a bit of data; However, the delayed effect of an information mining request is the study of the entire subject of the database. Information mining can be nominated as follows:

- Association rule mining or market basket analysis
- Classification and computation
- Cluster Analysis and Outdoor Analysis
- Web Data Mining and Search Engines'.
- Evolution Analysis

The proprietary of data mining comes into the habit of revealing the weak data that is undesirable by providing significant information mining. Figure 2 displays security information security tools for security, and data mining systems expose information that can benefit profitable information. Though systems are associated with it, it is not central to source information and can reveal important information.

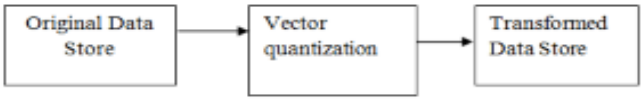

Fig 1: - Process of Vector Quantization

Some safety-security techniques may be associated with the ultimate goal of securing the security of the weak data, and then the mines.

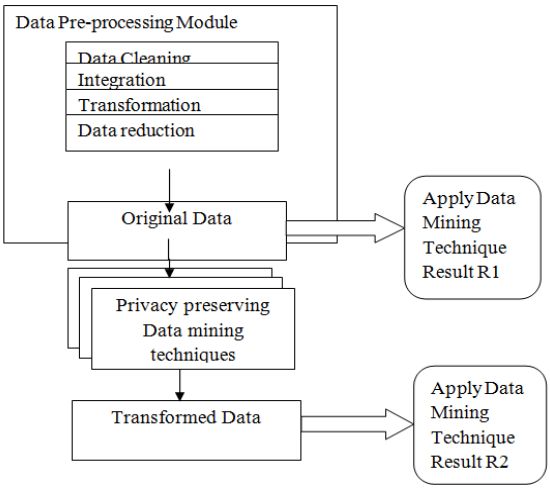

Fig 2: - Privacy Preserving Module

 DOI : 10.35291/2454-9150.2019.0226

Privacy preserving data mining has much significance because of following reasons:

1. Data mining causes social and ethical problem, because it reveals data which should requires Privacy?
2. Privacy preserving data mining provides security to private data against unauthorized access is a long term achievement for data mining security research community and for the government Agencies.
3. Hence, the security issue is one of the emerging area that became valuable research area in data mining.

3.1. Modified K-Mean Algorithm

3.1.1. Modified approach K-mean algorithm:

The K-mean method is popular clustering algorithm and its use in data mining, image separation, biometric and many other fields. This algorithm works well with small databases. We have proposed a way to work well with large scale databases on this paper. The algorithm of the modulated k is avoided to obtain at least some of the appropriate solutions at home and reduce the size of the cluster-error.

1. All data-points in the set D can be calculated by interval between each data point.
2. Find the closest pair of data points from the set
3. D and Data-Point system Amm (1 <= p <= k + 1), which contains two data points; Remove these two data points from the set D
4. To find the data point near the AP point set in the data point, add it to AP and delete it from D.
5. Repeat step 4 until AM reaches the number of data points (n / k).
6. If p + k + 1 is p = p + 1, another pair of D is to find data points, distance from the distance, Ap and point out another data-point system from D, step 4

Vector measurement (VQ) is the bulk used for information capture. In the past days, the structure's philosophy of a vector quiz (VQ) is regarded as an enormous complex on the need for a multi-dimensional mix.

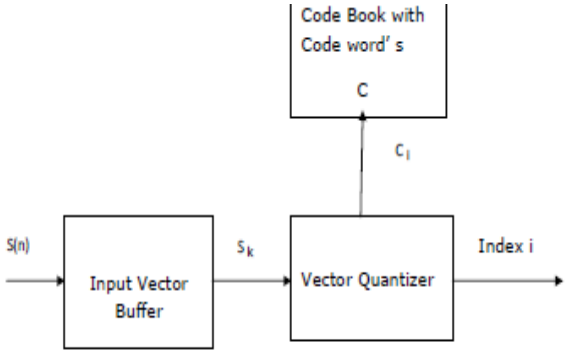

Figure 3: - VQ & Code Word generation

Heuristic-based techniques, encodes only selected values that minimize the utility loss rather than all available values.

Figure 4: - Code Book Preparation Process

It is reconstruction-based techniques where the original distribution of the data is reconstructed from codebook.

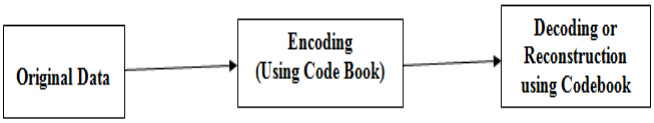

Fig 5: - Reconstruction of Code Word from Original Dataset

A specific factual purpose is to create a coded book, and each statistic (division) of the preparedness is reserved for w wards, each length l. These sections are collected using K to build the precedents for using the code. The code has centauries in all corners built by K. Each area of length L is approximated and is centered on length L, speaking Length L, which means the collection of the system, which closely focuses on the distribution size in the index book.

IV. PERFORMANCE EVALUATION

In this section, we evaluate our anonymous data integrity ethical performance and conduct experimenter analysis and experiments and provide excellent group instruction. A single test of this study is defined by a number of applications, number of mobile devices, sensor types in the sensitivity area, and the normal types of mobile devices.

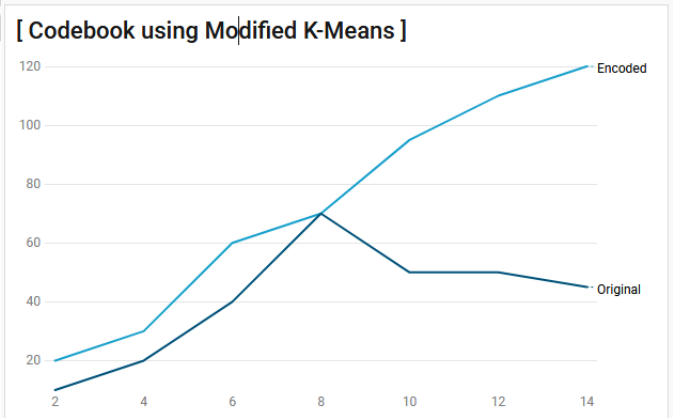

Fig 6: - Representation of original data and Encoded data

Privacy Preserving groups are an important architectural target for calculations and their productivity. In the above conclusions, the K-meaning LBG guidelines explain the performance data we use with the help of vector quality techniques.

 DOI : 10.35291/2454-9150.2019.0226

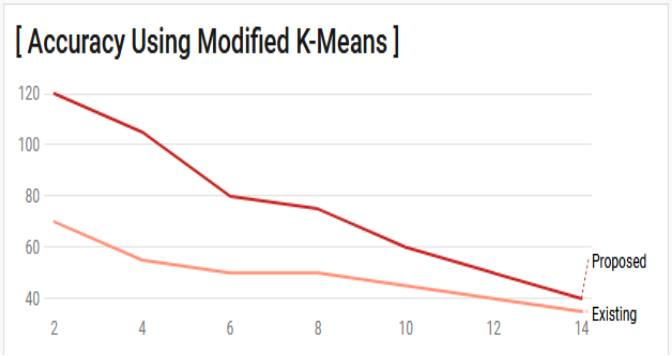

Fig 7 - Accuracy level

V. CONCLUSIONS

This job is engaged throughout vector quizzes; This is another method for insuring the insurance requirement by utilizing this index strategy. These lines can not reveal the main data for security. Meanwhile the right gathering can get results. The proposed program demonstrated how the control mechanism needed to ensure security can be a negative expression of how much of the mining can be expanded. Later, the planned program discovered another modified K-fish, especially with the reshaping of the film, expected to reduce these side effects. At this time we made clear and more relevant formulas to evaluate the execution, and we used the development method to detect the curve parameters to get better safety, accuracy, and possibility.

REFERENCES

[1] Johny Antony P, Dr. Antony Selvadoss Thanamani, "An Efficient Privacy Preservation Frame Work for Big Data using IRSA", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 11 (2018) pp. 8936-8940.

[2] K. Venkatesh, and Md. Ahamed, "Privacy Preserving Enriched Map Reduce for Hadoop Based Big Data Applications", National Conference Convergence of Emerging Technologies in Computer Science & Engineering.

[3] Antorweep Chakravorty, Tomasz Wlodarczyk, Chunming Rong, "Privacy Preserving Data Analytics for Smart Homes", 2013 IEEE Security and Privacy Workshops.

[4] Sarangkumar S. Dubey, Prof. Seema B. Rathod, "Implementation of Privacy Preserving Methods Using Hadoop Framework", International Research Journal of Engineering and Technology (IRJET).

[5] R.Sreedhar , D.Umamaheshwari, "Big-Data Processing With Privacy Preserving Map-Reduce Cloud", International Conference on Engineering Technology and Science-(ICETS'14).

[6] S.Sathya,Dr T.Sethukarasi, "Effective Privacy Preserving Method In Big Data Analytics For Healthcare Records", International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE).

[7] Nasrin Irshad Hussain, Bharadwaj Choudhury, Sandip Rakshit, "A Novel Method for Preserving Privacy in Big-Data Mining", International Journal of Computer Applications (0975 – 8887) Volume 103 – No 16, October 2014.

[8] Anju Abraham, Shyma Kareem, "Security and Clustering Of Big Data in Map Reduce Framework", International Journal of Advance Research, Ideas and Innovations in Technology.

[9] Kapilesh S. Swami1 , Dr. P Sai Kiran, "Security Methods for Privacy Preserving and Data Sharing Over Cloud Computing and Big Data Frameworks", International Journal of Science and Research (IJSR).

[10] Ashish P, Tejas S,Srinivasa J G, Sumeet G, Sunanda Dixit and Mahesh Belur, "Medical Application of Privacy Preservation by Big Data Analysis Using Hadoop Map Reduce Framework", International Journal On Advanced Computer Theory And Engineering (IJACTE).

[11] Anurag, Deepak Arora and Upendra Kumar, "Uml Based Model For Displaying Privacy Preserving Data Mining Systems", ARPN Journal of Engineering and Applied Sciences.

[12] Shalin Eliabeth S and Sarju S, "Bigdata Anonymization Using One Dimensional and Multidimensional Map Reduce Framework on Cloud", International Journal of Database Theory and Application.

[13] Ashish Lokhande, Sanjay Bansal, "Privacy Preserving Big Data Content Exposure Technique for Accidental Privacy Issues", International Journal of Contemporary Technology and Management.

[14] Mallappa Gurav, N. V. Karekar, Manjunath Suryavanshi, "Anonymization Of Data Using Mapreduce On Cloud", International Journal of Research in Engineering and Technology.

[15[Nandhini.P, "A Research on Big Data Analytics Security and Privacy in Cloud, Data Mining, Hadoop and Mapreduce", Int. Journal of Engineering Research and Application.

[16] Boel Nelson, Tomas Olovsson, "Security and Privacy for Big Data: A Systematic Literature Review", 2016 IEEE International Conference on Big Data (Big Data).

[17] Ms. Ch.Likitha Sravya, Mrs. G.V Rajya Lakshmi, "Privacy-Preserving Data Mining with Random decision tree framework", IOSR Journal of Computer Engineering (IOSR-JCE).

 DOI : 10.35291/2454-9150.2019.0226