ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 9, September 2022

An Optimized Approach for Privacy Preserving of Big Data using GDTM and Random Number Generators with GNN

[1] D.Kavitha, [2] Dr.T.Adilakshmi, [3] Dr.M.Chandra Mohan

[1] PH. D Scholar, JNTUH, Kukatpally, Hyderabad, India
[2] Professor & Head, Dept of CSE, Vasavi College of Engineering, Hyderabad, India
[3] Professor of CSE & Director of Evaluation, JNTUH, Computer Science & Engineering, Hyderabad, India
Email: [1] kavithadasari.it2005@gmail.com

Abstract— In this paper we propose a Privacy preserving mechanism of big data using GDTM along with Random Number generators. Given the rapid explosion of data being used across Enterprises, Individuals and Sensors, billions of data is being streamed and exchanged across the network. There is a high possibility of sensitive data being exchanged and stored, it's important to preserve sensitive data of Individuals and Enterprise data.

Most of the current techniques of privacy preserving in particular in the areas of data perturbation has been done on Static data. Given the dynamic nature of the applications and the huge data that is being generated it's important to evaluate the privacy preserving on big data without losing the accuracy.

Our research contribution is on Privacy preserving of big data using Geometric data transformation, random number generator and GNN techniques [5].

We would like to extend our research further on improving the accuracy of big data..

Index Terms— Privacy Preserving, Rotation, Translation and Scaling, GNN, Geometric Data perturbation, Random Number Generators

I. INTRODUCTION

Every day, vast amounts of data is being generated and collected by Enterprises, and devices(sensors) across the globe which might have details of customers, competitors etc. Numerous organizations and Enterprises analyses this data for crucial findings, decision makings and discoveries.

Data mining is predominantly used for extracting knowledge from huge amounts of data during this process there is a high probability to disclose sensitive information about individuals and organizations. Securing of sensitive data from unauthorized users when sharing information across the wire is of prime importance.

Privacy preserving of sensitive data has become a great concern, our research primarily addresses this concern by using Geometric data perturbation techniques, random number generators and GNN on big data.

Our Contribution can be summarized as follows:

1. For the first time we have applied the Geometric data perturbation (Rotation, Translation) with random number generators on the big data using GNN and observed high accuracy [3].
2. As a next step, to further improve the accuracy, we have added scaling perturbation technique to the above-mentioned methodology.

II. RELATED WORKS

Merve Kanmaz[14] proposed a methodology is based on Geometric data perturbation and generating of noise using random numbers. This method made use of Linear Congruential Generator for generation of the noise. This method has shown better results compared with existing data perturbation techniques. The research was conducted using 5 different data sets and evaluated using NB, J48 and DT and OneR. The methodology was evaluated again Attack Resistance using ICA and has shown promising results.

Aldeen, Y.A.A.S., Salleh, M. & Razzaque, M.A. proposed PPDM techniques based on distortion, associative classification, randomization, distribution, and k-anonymization. In this model sensitive data has been masked, transformed and retrieve the transformed outcome.

Chen, K., Liu, L. proposed Geometric Data Perturbation (GDP) method in this model which selectively preserves the task/model specific information in perturbation which has shown better privacy guarantee. GDP method was compared against well-known data-mining models and random projection perturbation and also evaluated against different level of attacks and proven. It has been proved GDP has provided better privacy guarantee and good accuracy as well.

F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner and G. Monfardini, proposed on GNN. A learning algorithm to estimate model parameters was provided and demonstrated that the method is suitable also for large data sets. Based on the experimental results, solid results were observed.

III. DATASET DATASET DETAILS

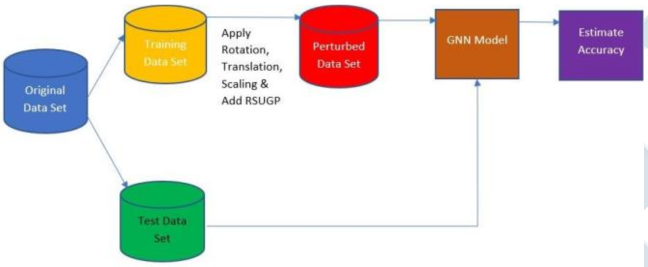
Traffic Based IoT Sensor Big Data has been used. http://iot.ee.surrey.ac.uk:8080/datasets.html

Base Dataset Shape (3795475, 9)

The test data is split into the ratio of 75(Training):25(Testing).

IV. PROPOSED METHODOLOGY

Geometric data perturbation (Rotation, Translation and Scaling) with random number generator (Noise) on big data using GNN. Below figure shows details of the proposed method.



4.1 Geometric Perturbation

One of the widely used perturbation technique is Geometric Data Perturbation.

Data is rotated using a random angle, translated, scaled and finally noise is added to the data (calculated using random generators) based on below formula.

RX represents Rotation, T represents Translation, S represents Scaling of the original data in sequence,

Δ-Adding of Noise to the data:

$$G(X) = RX + T + S + \Delta$$

4.2 Noise addition

Noise is generated using **Linear Congruence Generators**

$X_{i+1} = aX_i + c \pmod m$

'm','a' and 'c' represents the maximum value, standard deviation and average values in the data.

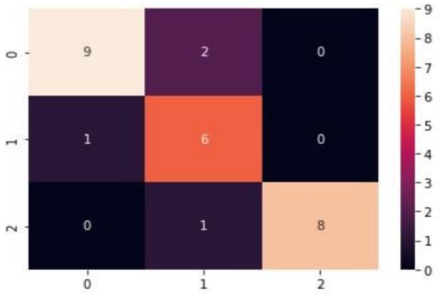
V. RESULTS AND DISCUSSION

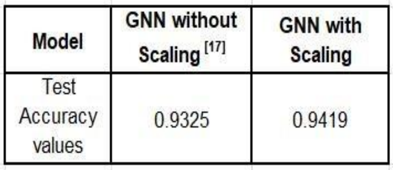
5.1 Classification Accuracy

Classifier model performance has been evaluated using below formula.

$$\text{Classification Accuracy} = (T\,P + T\,N)/(P + N)$$

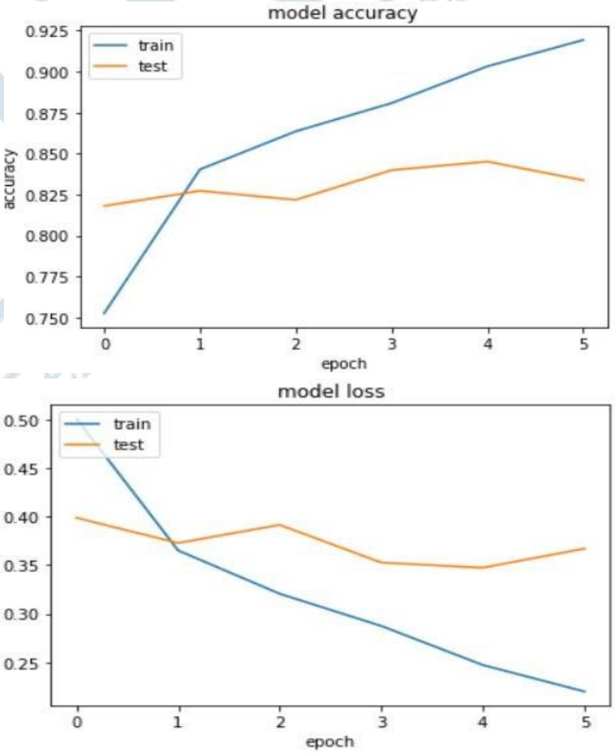
Confusion Matrix &Accuracy Comparison of GNN with and without Scaling





Model	GNN without Scaling [17]	GNN with Scaling
Test Accuracy values	0.9325	0.9419

5.2 Graph showing Accuracy vs Loss in GNN Training



VI. CONCLUSION

As the volume of data increases particularly big data, it's important to preserve and retrieve the data correctly. With the proposed approach training and classification accuracy has increased by adding scaling transformation to Geometric data perturbation with random number generators using GNN [15]. Proposed method has proven to be a feasible approach for Privacy protecting of sensitive data and Classification Accuracy has increased from 93% to 94%.

REFERENCES

[1] Chen, K., Liu, L. Geometric data perturbation for privacy preserving outsourced data mining.Knowl Inf Syst 29, 657–695 (2011). https://doi.org/10.1007/s10115-010-0362-4

[2] Aldeen, Y.A.A.S., Salleh, M. & Razzaque, M.A. A comprehensive review on privacy preserving data mining. SpringerPlus 4, 694 (2015). https://doi.org/10.1186/s40064-015-1481-x

[3] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner and G. Monfardini, "The Graph Neural Network Model," in IEEE Transactions on Neural Networks, vol. 20, no. 1, pp. 61-80, Jan. 2009, doi: 10.1109/TNN.2008.2005605.

[4] Merve Kanmaz1,∗ , Muhammed Ali Aydin2 and Ahmet ―A New Geometric Data Perturbation Method for Data Anonymization Based on Random Number Generators‖ https://journals.riverpublishers.com/index.php/JWE/article/view/7983.

[5] D. Kavitha, Dr. T. Adilaxmi, Dr. M. Chandra Mohan. (2022). Efficient Privacy Preservation of Big Data Using Random Number Generators and Geometric Data Transformations. Mathematical Statistician and Engineering Applications, 71(3), 268 –. Retrieved from https://www.philstat.org.ph/index.php/MSEA/article/view/164

[6] https://www.sciencedirect.com/science/article/abs/pii/S0167739X17324391?via%3Dihub