International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

A Sigmoid based Learning in Heterogeneous Distortion for Data Privacy

K Sandhya Rani Kundra, J. Hyma, P.V.G.D Reddy, K. Venkata Rao

Abstract: ***Neural network-based learning models along with an access to huge data have made a remarkable outcome in recent years. These models are contributing a lot to improvise the working dimensions of various domains like Speech recognition, Image processing, Text analysis and many more. The well represented data is the main resource in the current research, but this data is often privacy sensitive and it definitely needs a proper attention failing which leads to serious privacy concerns. The proposed work demonstrates how learning models can be applied to analyze the data sensitivity and classify them to various privacy classes. Once the privacy class distribution is performed the model applies Inverse laplacian query model to check the data utility. The data should not get compromised on utility with the curse of privacy. With this intention the given experimental study succeeded in training the network to perform privacy analysis under a modest privacy budget, complexity training efficiency and data utility.***

Keywords: ***Neural Networks, Differential Privacy, Query Model, Data utility.***

I. INTRODUCTION

The availability of the large and representative data sets facilitates the neural network models to come up with their progression in a broad range of applications like Image classification, language processing, data analytics etc [1]. These rich informative datasets collected by powerful sensors and high mobility devices are the major source for many of the state of art techniques in current research. The learning models when exposed to such data, results in deeper analysis and could be able to extract useful predictions [2, 3, 4, 5]. However, this data is often crowd sourced and may contain private information of the individuals. So, the privacy guaranteed techniques while offering the application requirements are in much demand. The earlier work [6, 7] explored the significance of privacy preservation while performing deep learning analysis. Very less study has been noticed in the context of applying the deep learning algorithms to perform the privacy analysis. As privacy parameter always argued to be heterogeneous in nature, to achieve this, data needs to be classified into various classes according to the requirements. Various constraints for data mapping into different privacy classes and their utility are verified with the query model in the work proposed in [8,9].

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
K.K.Sandhya Rani ,Asst.Professor, Department of Information Technology, Gayatri Vidhya Parishad College of Engineering(A), Andhra Pradesh, India
Dr. J.Hyma, Associate Professor, Department of Computer Science Engineering, ANITS, Andhra Pradesh, India
Prof P.V.G.D Reddy, Vice Chancellor, Sr.Professor of Computer Science & Systems Engineering Department, Andhra University, Andhra Pradesh, India
Prof.K.Venkata Rao, Academic Dean and Professor, Computer Science & Systems Engineering Department, Andhra University, Andhra Pradesh, India

In this work a new algorithmic technique is developed with a refined privacy analysis that could work within the framework of differential privacy-based query model. This technique automates the constraint specification [10] using a learning model to map the data items into various privacy classes without human intervention. The data along with its privacy class distribution is then fed as an input to the query model. The query model will work on the basis of inverse laplacian distribution and it more detailed in further sections.

II. RELATED WORK

Privacy preserving aspects in big data modelling are demonstrated in the work proposed in [11]. A deep learning based computational model has been studied with operations in cloud environment. However, when dealing such huge data in the cloud, privacy protection is one of the major challenges. A sigmoid function is used as a polynomial function to find the approximations in the process of secure computation. Their experimental study proved that their model is highly scalable. Another work proposed in [12] implemented and evaluated a practical system that enables two or more parties to learn their objective without disclosing input data sets. The given experimental study was carried out with a differential privacy algorithm with different 'ε' values and it was practically shown that smaller the 'ε' values greater the privacy protection and results in lower accuracy. Another experimental study on deep learning-based privacy preservation along with differential privacy is exemplified in [13]. Two standard image classification techniques MNIST and CIFAR-10 are proposed and their evaluation mentioned that the privacy protection in deep learning analysis is very much required. Automatically identifying sensitive objects in images with convolutionary neural networks has been studied in [14]. Their work aimed at detecting sensitive objects in the images and there by recognizing their privacy class to automatically recommend the necessary privacy settings while sharing the images in any social media. The work proposed in [15] has given a study over two activation functions namely Sigmoid-weighted Linear Unit (SiLU) and its derivative function (dSiLU) for required approximations in learning. Their study demonstrated how effectively these activation functions are used for reinforcement learning. The challenges in applying the deep learning techniques over the encrypted data stored in a third-party cloud have been studied in the work presented in [16]. Their work proposed multi key fully homomorphic encryption (FHE) for performing a learning objective without disclosing the input data to the parties in the privacy preserving manner. Another work proposed in [17] attempted to propose a privacy preserving deep learning system to learn over a combined data set without disclosing the data to the other participants.

Their work used additive homomorphic encryption to achieve the desired privacy goal. A new variation of differential privacy called Matrix Variate Gaussian (MVG) mechanism is proposed in [18]. Their work explored how (ε,δ) differential privacy technique is preserving the privacy component over matrix valued queries, and also studied impact of noise on the data utility.

A. Our contributions and Implementation strategies:

The existing literature motivated to deliberate the different working aspects of learning models in obtaining the heterogeneous privacy protection. Major contributions include

1. At the first instance, the detailed impact of privacy loss is pursued with the wider literature study. This enabled much more stringent estimates of the overall privacy component.
2. For computing gradients of each individual training samples various heterogeneous privacy requirements, correlation constraints and domain knowledge constraints proposed in our earlier work [8,9,10], are extended in this framework.
3. A network is built to map each sample to its corresponding privacy class.
4. This learning framework along with differential privacy has been evaluated against a standard data set and verified for privacy and data utility.

III. EXPERIMENTAL SETUP

The evaluation of the proposed model is twofold. Firstly,it applies a sigmoid activation function which will let the network to learn and obtain the privacy class of each data sample. One the privacy class either high or low is obtained then it is evaluated with query model drawn from the basic idea of differential privacy.

A. Sigmoid Function

Sigmoid functions are the building blocks of deep neural network, unlike other neural network functions sigmoid functions give a better hand dealing with the non-linear data by providing a continuous output between 0 to 1 as a probability range where as other neural network functions like perceptron gives a step function as output which has a huge disadvantage while dealing with the non- linear data. Sigmoid function output is an S shaped curve which is much smoother than the step functions in the perceptron neural network, in perceptron for every small change the result might be a complete flip whereas in sigmoid due to its s shaped output, the transaction is much more smooth and for every small change the result might not change drastically.

Input:

The input to sigmoid are real numbers and the output will be in the range of 0 to 1, where by choosing a threshold can be used to classify the results into binary.

Step-1: Initialise the parameters w, b

Step-2: Iterate until satisfied

Compute L (w, b)

$$w(t+1) = wt - \eta \Delta wt$$

$$b(t+1) = bt - \eta \Delta bt$$

Here w and b are initialized randomly and iterate through the data, on every iteration the squared error is computed, depending on the squared error the parameters are updated such a way that the squared error is minimized.

$L(w,b) > L(w + \eta\Delta w, b + \eta\Delta b)$

The loss function is defined as follows

$Loss = \sum_i (Z_i - \hat{Z}_i)^2$

Where $\hat{Z} = \frac{1}{1+e^{-(wx+b)}}$

In sigmoid the main purpose is to update the parameters w and b so that the overall loss function of the model is reduced.

B. Inverse Laplacian Distribution

In Inverse-Laplace mechanism the noise parameter is drawn from an Inverse-Laplace distribution, which can also be expressed by probability density function [19, 20] given in equation 1.

$$\boldsymbol{noise(y) \propto exp\ (-|y|/\lambda\ )}\text{.....Equation 1}$$

With mean zero and standard deviation $\boldsymbol{\lambda}$. The output function of $\boldsymbol{A}$ is equation 2.

$$\boldsymbol{T_A(x) = f(x) + Y}\text{.....Equation 2}$$

Where $\boldsymbol{Y \sim Lap^{-1}(\lambda)}$ and f is the query or function posed on the database. Now clearly $\boldsymbol{T_A(x)}$ can be considered to be a continuous random variable [21] [22][23]. $\frac{\Delta(f)}{\lambda}$ being the privacy factor ϵ which is at most $e^{\frac{|f(D1)-f(D2)|}{\lambda}} \leq e^{\frac{\Delta(f)}{\lambda}}$. It is a derived fact that in order to have A as the ϵ - differential private algorithm we need to have $\lambda = \frac{1}{\epsilon}$.

IV. PROPOSED METHODOLOGY

Initially a part of data set is used to train the sigmoid network and the output of sigmoid network is in the range of 0-1, here the model sets a threshold i.e., if the output is between 0 – 0.5 then it is classified as Low privacy class and if it is in the range of 0.5-1 then it is classified as High privacy class. This classification results in the data along with its respective privacy class distribution. This step overcomes the disadvantage of treating all the data at the same uniform level in privacy protection. When a user gives a query then the corresponding results are extracted and subsets of different privacy classes are computed. According to the privacy class of subsets the noise is added using the inverse laplacian differential privacy technique. Finally, all the subsets are combined and results are featured.

The figure [1] depicts the flow of the system where the dataset is given to the sigmoid neural network which classifies the data set into low and high private classes based on all possible instances.

The algorithm Sigmoid-Hetro Data Distortion (SHDD) is given in section 4.1. The dataset D is given as input to the neural network which in turn applies the learning mechanism to obtain the D'. The D', output of the network will contain privacy mapping class High or Low. When the query is posed on this data heterogeneous distortion levels will be introduced and accordingly the modification will take place.

Retrieval Number: K24170981119/19©BEIESP
DOI: 10.35940/ijitee.K2417.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The privacy budget parameter will vary according to the class and finally contributes to the calculation of the noise parameter N. The modified answer with a balanced weightage of privacy and utility will be given to the user.

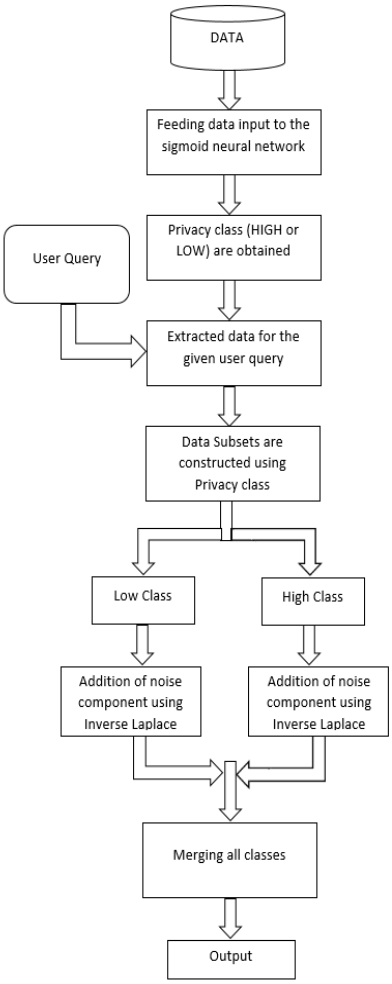

Figure [1]. System Architecture

A. Algorithm

Algorithm: Sigmoid-Hetro Data Distortion (SHDD)
Input: *Dataset D, Query function Q ()*
Output: *Perturbed Answer Q (D'')*
Start:
Step-1: *Dataset given input to sigmoid neural network which gives D'*
Step-2: *Based on User query the subsets are divided based on class factor*
Step 3: *Calculate the Noise Parameter* N_i
$N_i = Lap^{-1} (\Delta F_i / \epsilon_i)$ *//where* $i \epsilon$ *high||low*
Step 3.1: *For each data attribute* $d_i \in D$
If (class (d_i) *== high)*
Then $N_h = Lap^{-1} (\Delta F_h / \epsilon_i)$ *//for all* $i \epsilon$ *high*
Else
$N_l = Lap^{-1} (\Delta F_l / \epsilon_i)$ *//for all* $i \epsilon$ *low*
Step 3: $Q(D'') = Q(D') + N_i$ *//* $\forall i \epsilon$ *high||low*
Stop;

In previous work [8] the model worked on heterogeneous data where it followed the three-question model, which consists choices of user, database admin and co-relation factor this way our data was built on a huge scale. In order to use the that historical data efficiently the same has been used to train our sigmoid neural network which will be more efficient than our previous models, since it is an automated process it takes every attribute into account while computing the privacy class (HIGH or LOW) of the data.

V. RESULTS AND DISCUSSION

For perceptron model input is real but we normalise the data to make training faster and reduce the chances of stuck in local optima. Each attribute has an associated weight so that every attribute's importance can be in included in making decision. Initially weights are set to ones and b is randomly initialised. Weights are adjusted based on the difference between the threshold and dot product of weight and input vector. Epochs and learning rate help in better accuracy levels.

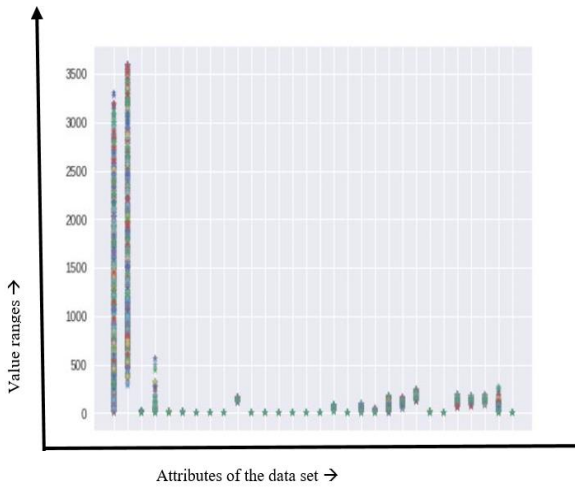

Figure 2: Value Ranges of Data set

The above fig depicts the value ranges of dataset, on x-axis the attributes are list and on the y-axis the attributes value range is shown. It clearly shows the values are not binary in nature; the sigmoid function can deal with variable ranges of data.

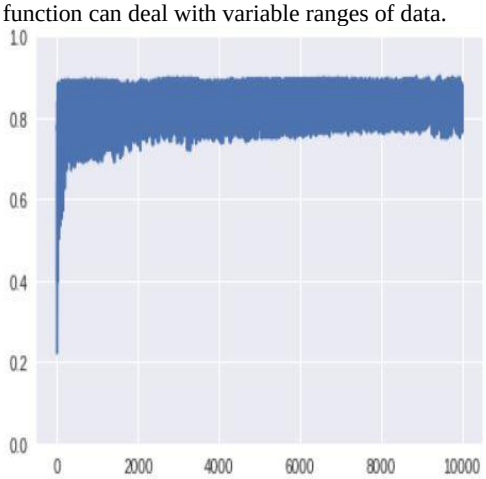

Figure 3: Weight adjustments of Sigmoid

The above figure [3] depicts the weight adjustments on every iteration of the model; here initially weights are adjusted randomly on every iteration the loss function is computed based on the loss function weights are manipulated accordingly.

Database Income Attributes present: 9 [Salary, Commision, Age, Elevel, Car, Zipcode, Hvalue, Hyears, Loan, Class]

No of records: 32519

Query1: What is the average salary for people without commission (i.e.commission = 0)

Query2: What is the maximum salary for people in the age limit 40 to 60

Query3: What is the minimum salary for people who have only one car

Query4: What is the avg loan for people living in zipcode 530009

Query5: What is the sum of havalues where hyears is greater than 20

Query6: What is the average age of people without loan (i.e. loan=0)

Query7: What is the maximum no of cars owned by an individual

Query8: What is the average Elevel

Table [1] : Query Evaluation with various privacy techniques

	Original	Homo-Diff Privacy	Hetro-Diff. Privacy	Sigmoid hetro-Diff. Privacy
Query 1	146.91	147.92	149.92	148.745
Query 2	15.37	16.37	17.8731	16.985
Query 3	407.05	408.05	409.5521	409.23
Query 4	0.04	1.17	3.3018	2.986
Query 5	26.07	27.07	28.573	27.856
Query 6	66760.79	69279.49	70781.2	70753.21
Query 7	23213.13	23419.68	23565.73	23543.55
Query 8	5.69	6.75	8.85	7.99
Mean	11321.88	11663.31	11870.63	11863.819
Mean Absolute error	nil	341.43	548.74	541.939

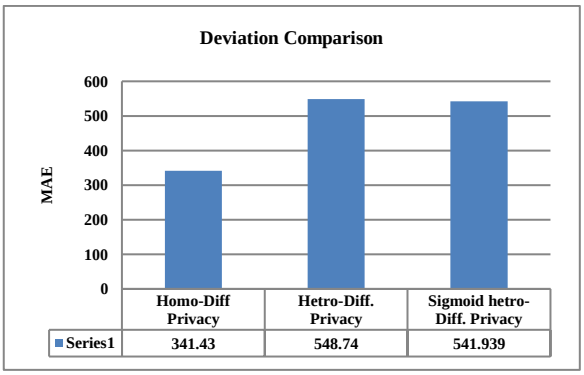

Figure 4: Result comparisons

The results of homogenous, heterogeneous and sigmoid – heterogeneous Differential privacy are given in table [1]. From the above table it is evident that there is noticeable amount of deviation when compared with each other. In homogenous differential privacy the utility rate is very high but the data privacy is very low in order to improve the privacy a better technique is used, in this case the model used Heterogeneous differential privacy where there is a good balance of privacy and utility but the utility rate is low when compared to the homogenous differential privacy. In order to improve the aspect of utility and also not to compromise on the privacy at the same time, a new learning model namely Sigmoid Heterogeneous differential privacy technique is imposed. It provides a better utility rate with the same amount of privacy offered as the Heterogeneous Differential privacy with less human intervention.

In all our previous works we have used data and imposed rules in order to make a heterogeneous privacy models, the major drawback is that every time we have define the rules for every new dataset. In order to overcome that, here all the historical data with the rules imposed are given as training to the sigmoid algorithm, which in turn gives the privacy label of each instance as high or low based on all the attributes present.

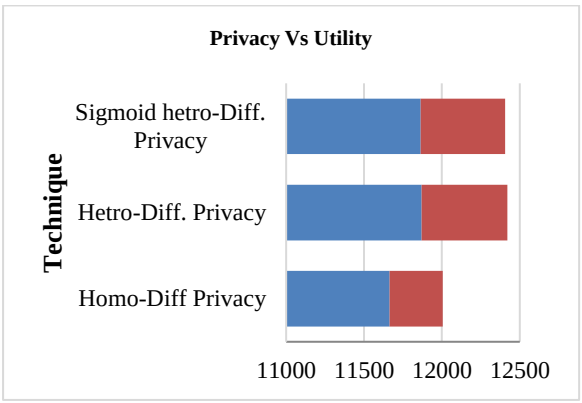

Figure 5: Privacy Vs Utility

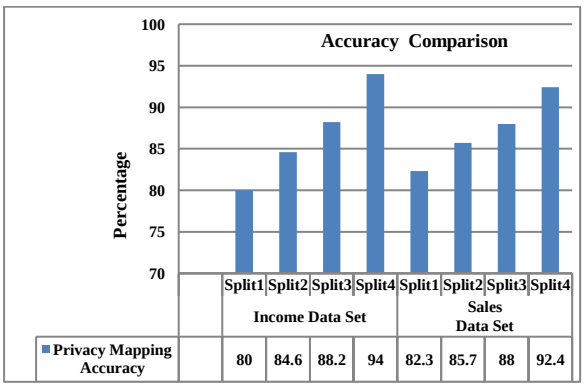

Figure 6: Hetro Vs Sigmoid Hetro Differential Privacy

The above depicts the privacy and utility levels of each technique, where Homo-Diff Privacy has high utility rates but low privacy rates, in hetro – diff privacy and Sigmoid hetro – diff privacy the rates are very close but the advantage of using sigmoid differential privacy technique is it gives more flexibility to adapt to the techniques i.e., all the backend process of construction of 3 party question model is not needed like it is in Hetro – diff privacy technique because sigmoid hetro diff privacy uses historical data and automates the process which will be more efficient than the previously used techniques.

Retrieval Number: K24170981119/19©BEIESP
DOI: 10.35940/ijitee.K2417.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

VI. CONCLUSION

The work proposed in this paper aimed to collectively use a sample learning technique for training a network to classify the given input data into various privacy classes. Sigmoid activation function is successfully derived to approximate the privacy class of the each individual data item. When the user poses a query the targeted data will be modified according to its privacy class. Different 'ϵ' value is mapped for different privacy classes and this leads to heterogeneous data distortion rather than uniform distortion. This heterogeneously distorted data now will become the input to the user query. The query result is observed to have controlled amount of noise. This is because of the nature of differential privacy distribution. Though there is no much amount of deviation between Hetro –Diff privacy and Sigmoid based technique, definitely the network based model is better useful in predicting the privacy class in order to achieve the heterogeneity. Once the model is well trained, it is observed that it could predict the class of the other set of input data with good amount of accuracy. The work can be further extended with advanced learning models to accommodate more number of classes and it can be verified against various types of data.

REFERENCES

1. Y. LeCun, L. Bottou, G. Orr, and K. Muller. Efficient backprop. Neural Networks: Tricks of the trade, Springer, 1998.
2. G.F. Montufar, R. Pascanu, K. Cho, and Y. Bengio. On the number of linear regions of deep neural networks. Advances in Neural Information Processing Systems, 27:2924–2932, 2014.
3. R.M. Neal. Bayesian learning for neural networks. Springer Science & Business Media, 118, 1995.
4. R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. Proceedings of the 30th International Conference on Machine Learning, 28:1310–1318, 2013.
5. B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein, and S. Ganguli. Exponential expressivity in deep neural networks through transient chaos. 30th Conference on Neural Information Processing Systems, 2016.
6. White House Report. Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy. Journal of Privacy and Confidentiality, 2013.
7. John Duchi, Michael I. Jordan, and Martin J. Wainwright. Privacy aware learning. Journal of the Association for Computing Machinery, 2014.
8. K Sandhya Rani Kundra, J Hyma, PVGD Prasad Reddy and K Venkara Rao Privacy preserving query model using inverse laplacian differential technique,Published under licence by IOP Publishing Ltd Journal of Physics: Conference Series, Volume 1228, conference 1
9. Mohammad Alaggan, ebastien Gambs and Anne-Marie Kermarrec 2016 Heterogeneous Di fferential Privacy Journal of Privacy and Confidentiality 7 p 127-58.
10. J.Hyma, PVGD Prasad Reddy, A.Damodaram "A Study of Correlation Impact on Privacy Preserving Data Mining" ,IJCA, Vol 129, 2015.
11. 8.Q. Zhang, L. T. Yang, and Z. Chen. Privacy preserving deep computation model on cloud for big data feature learning. IEEE Trans. Computers, 65(5):1351–1362, 2016.
12. 4.Reza Shokri and Vitaly Shmatikov. Privacy-preserving deep learning. In Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security, CCS '15, 2015.
13. Martin Abadi, Andy Chu, Ian Goodfellow, Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In 23rd ACM Conference on Computer and Communications Security (ACM CCS), 2016.
14. Jun Yu, Baopeng Zhang, Zhenzhong Kuang, Dan Lin, Jianping Fan iPrivacy: Image Privacy Protection by Identifying Sensitive Objects via Deep Multi-Task Learning. IEEE transactions on information forensics and security, 2016.
15. S. Elfwing, E. Uchibe, and K. Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. arXiv:1702.03118, 2017.
16. P. Li, J. Li, Z. Huang, T. Li, C. Gao, X. Liu, K. Chen, "Multi-key privacy-preserving deep learning in cloud computing", Future Gener. Comput. Syst., vol. 74, pp. 76-85, 2017.
17. Aono, Y., Hayashi, T., Wang, L., Moriai, S., et al. (2017). Privacy-preserving deep learning via additively homomorphic encryption. IEEE Transactions on Information Forensics and Security.
18. Thee Chanyaswad, Alex Dytso, H Vincent Poor, and Prateek Mittal. Mvg mechanism: Differential privacy under matrixvalued query. In Proceedings of the 25nd ACM SIGSAC Conference on Computer and Communications Security. ACM, 2018.
19. Dwork C, McSherry F, Nissim K and Smith A 2006 Calibrating Noise to Sensitivity in Private Data Analysis.
20. Rathindra Sarathy and Krish Muralidhar 2009 Differential privacy for numeric data in proceedings Joint UNECE/Eurostat work session on statistical data confidentiality.
21. Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science. Now Publishers, 2014.
22. Rathindra Sarathy and Krish Muralidhar 2009 Differential privacy for numeric data in proceedings Joint UNECE/Eurostat work session on statistical data confidentiality.
23. Zeyu Ding, Yuxin Wang, Guanhong Wang, Danfeng Zhang, and Daniel Kifer. Detecting violations of differential privacy. In ACM Conference on Computer and Communications Security, 2018.

AUTHORS PROFILE

Mrs K.K.Sandhya Rani ,Asst.Professor in the department of Information Technology,Gayatri Vidhya Parishad College Of Engineering(A),currently perusing Ph.D. in Andhra University,Visakhapatnam. Interested areas are Privacy and security and information security.

Dr.J.Hyma,Associate Professor in the department of computer science engineering, ANITS. Her area of interest in Data Science, Internet of Things, Privacy and Security.

Prof P.V.G.D Reddy is presently Vice Chancellor, ANDHRA UNIVERSITY and Sr.Professor of Computer Science & Systems Engineering department which is the largest department in entire South India, and also serving as MEMBER, Executive Council of VIGNAN Deemed University, Guntur. He has been awarded the Best Teacher Award for the year 2011 by the Govt. of Andhra Pradesh in the combined state. Prof. Reddy's Research areas include Soft Computing, Software Architectures, knowledge Discovery from Databases, Image Processing, Number theory & Cryptosystems. He has 3 Patents granted.

Prof.K.Venkata Rao, presently Academic Dean, ANDHRA UNIVERSITY, and Professor of Computer Science & Systems Engineering department. He is currently director & chairman; teachers mutually aided cooperative society ltd. and. Honorary director and web master, Andhra University.