Implementation of Privacy Preserving Data Classification System Using Machine Learning Technique

Miss. Sayali B. Desale, Prof. S.B. Javheri
Department of Computer Engineering Rajarshi Shahu College of Engineering, Pune, India.

Abstract— **A decision support system is a computerised process which can help to make decision and action with users and help users to gain the extra knowledge. In this system server has rich dataset for particular application. In this scenario, data arranged in outside servers through the web by data owners for analysing user's results. But these servers are outsider servers which are not completely trusted, so here comes a privacy concern. The main purpose of this system is to design a novel privacy-preserving data classification system by using the machine learning technique Artificial Neural Network for the purpose of classification. To provide a security we use Paillier homomorphic encryption technique where all the computations can be done in encrypted form. Neither the server involved in the process nor the user ready to realize any additional learning about the server side parameters and results. So, user's data as well as categorization parameters from server side will secure and the security of user's information is not bargained.**

Index Terms— **Artificial Neural Network, back-propagation , Homomorphic encryption, privacy preservation.**

I. INTRODUCTION

These days outsider outsourcing gives efficient and exact decision support in numerous applications. In this situation, data owner can use the data available in outside servers through the web to analyse the result of the user's data. But these remote servers are third party servers which are semi trusted raises the security and protection concerns. In this framework, a novel protection saving data classification system is being form where the user's data dependably stay in an encoded structure amid the conclusion process. As a result, the server included into the conclusion procedure is not ready to catch additional information about the user information and results. In this system, a decision support system can be built at server side using the existing datasets. Now user's able to analyse their results rely on their information by sending the users data to the server via the web to perform operations taking into account the learning at the server side. It reduces the burden of user to construct his self-decision support system which require gathering of bunch of data and storage resources.

At the server side system uses the Artificial Neural Network, which is the machine learning technique for the sake of classification of given information. We use Back propagation scheme for neural network learning. We use the paillier homomorphic encryption scheme where all the operations are carried out in scrambled form which saves the privacy of user's data as well as categorization parameters from server side.

II. LITERATURE REVIEW

Yogachandran Rahulamathavan [1] has proposed protection safe guarding choice supportive system for medical field for medical diagnosis process. In this paper they form protection safe guarding choice supportive system at server side by making the use of existing available datasets from UCI database. In this paper they use Support vector machine for the sake of classification .To provide security they use paillier homomorphic encryption technique by using Gaussian kernel based categorization which provides security to patients data and server side choice making parameters.

D. J. Bonde [2] has proposed number of survey procedures of information protection in cloud using back spread neural system. In this paper various gatherings perform collaborative learning on arbitrarily partitioned information by utilizing distributed computing. In which for protection safeguarding every gathering send plain content to the cloud and cloud scramble that content. So that the computation and communication cost will be diminish.

Mr. Anwar Basha. H [3] has proposed protection safe guarding back-spread neural system learning utilizing cloud computing. he learning as a part of neural system happens with the assistance of Back Propagation Algorithm. In this paper the cloud stage is utilized for the learning procedure to happen in a protected manner. The various gatherings take an interest and share their datasets to make the learning process through back propagation successfully. The Trust Agent circulates the keys in the current framework. The cloud conveys the keys in the proposed framework. This is a change subsequent to the cloud does not know the private information of the proprietors since they are encoded first and after that transferred in cloud.

Arso M. Vukicevic [4] proposed a choice supportive network for medical field for the evaluation of bone force intensity variable by utilizing artificial neural systems (ANN). The technique catches the stress intensity variable as indicated by patient's age and analyzed split length. ANN was prepared utilizing the test information accessible as a part of literature. They got results demonstrated great relationship with the trial information, with potential for further enhancements and applications.

Mrudula Gudadhe [5] proposed a choice making supportive system for medical field for heart disease. This paper shows a choice emotionally supportive network for heart illness grouping in view of support vector machine (SVM) and Artificial Neural Network (ANN). A multilayer perceptron neural system with three layers is utilized to build up a choice emotionally supportive network for the analysis of coronary illness. The multilayer perceptron neural system is prepared by back-propagation calculation which is computationally productive strategy. Results got demonstrate that a MLPNN with back-propagation can be effectively utilized for diagnosing coronary illness than support vector machine.

R. R. Janghel [6] proposed a clinical choice supportive network by utilizing the obsessive ascribes to decide whether the fetal delivery to be done ordinary or by surgical technique. The neurotic tests like glucose, pulse (BP), resistivity list (RI) and systolic-diastolic (S/P) proportion will be recorded at the season of delivery. All characteristics exist in a particular reach for typical patient. The database comprises of the qualities for cases i.e. typical and surgical conveyance. Here they used artificial neural network for test system. Three models of ANN are prepared utilizing backpropagation calculation (BPA), radial basis function network (RBFN) and learning vector quantization network (LVQN). This framework will help specialist to take choice at the basic time of fetal conveyance.

III. IMPLEMENTATION DETAILS

A. System Overview

The system work as follows:

- First data owner can form choice making supportive system by using existing available dataset from UCI-database. Here we are using IRIS-Plant dataset which having 4 attributes, 3 classes and 150 instances.
- User use Paillier homomorphic encryption to do secure computation. First user having his own public key and private key. Then user can send the data in encrypted form a to server. User encrypts his data by using his public key. User can keep his private key secret. Users data must be in a normalize form.
- At the server side data classification can be done by using artificial neural network. At the server side we are using Backpropagation alogorithm to train the network. After neural network gets train we are doing actual testing.
- We use the paillier cryptosystem to perform operations in encrypted domain. The server can perform all computations on encrypted data and sends result in an encrypted form to user. Only user able to decrypt the result by using his private key.
- As server only has encoded users information. So, it is necessary to compute result in encrypted domain.
- Then result is calculated and decision is send to user which is in the encrypted format .Then user can decode it by using his private key.
- Overall computation between user and server can be done in encrypted form. User as well as server will not able to extract any information of each other. So, user's data as well as categorization parameters from server side will secure.

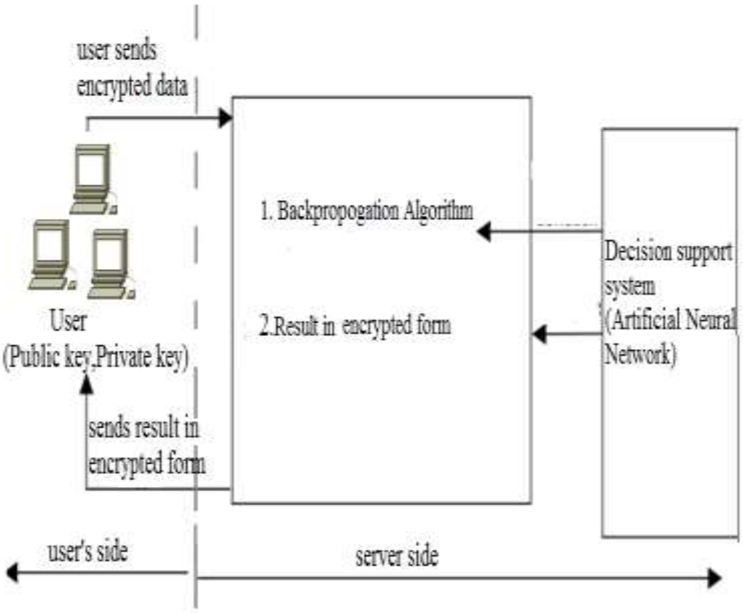
user sends
encrypted data
1. Backpropogation Algorithm
2.Result in encrypted form
Decision support system (Artificial Neural Network)
User
(Public key,Private key)
sends result in encrypted form
user's side
server side

Fig. System Architecture

B. Decision Support System

It is a computerised procedure which can help to make related decisions and activity with users and to pick up the additional learning. User manually enters the characteristics or retrieve it from existing records. Here decision support system provides numerous supports like alert of critical issues, preventions, suggestions etc.

C. Artificial Neural Network

Artificial neural network (ANN) is a machine learning approach that models carnal brain and consists of a number of Stirred neurons. Many times neuron in ANN receives a number of inputs. An activation play the part is practical to these inputs which modify gives the output value of the neuron. An Artificial Neural Reticule consist of neuron incise which processes the intimate of the NN, an origination which consists of belt of neurons and links which connects that neurons.

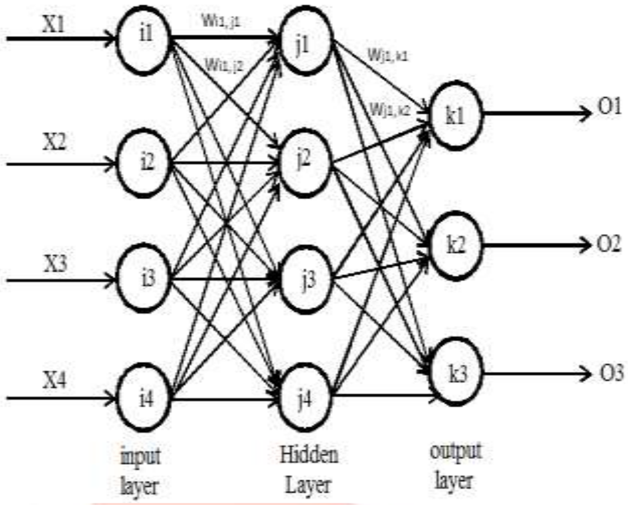

Fig. Neural Network

Every neuron inside the network framework is basically a simple processing unit which takes one or more inputs and generates an output.

Artificial neural network consist of two phases namely training phase and testing phase. In training phase we can train the network by using some instances from existing dataset we are taken .Here dataset having 150 instances. So out of 150 instances we are using 120 instances for training purpose. After network will be train we done the actual testing by giving other remaining 30 instances as input.

At each neuron, every input has an associated weight which updates the strength of each input. The neuron make addition of all inputs and calculates an output to be passed on. Here we are using back propagation training algorithm which adjusts the weights of neural network to minimize the total error of the network over training set.

D. Paillier Homomorphic Encryption

The main base of the protection safeguarding decision support framework is Paillier Homomorphic encryption. It is a public key encryption technology. It is the encryption form where all the computations are done in cipher text only. In this system we are going to use paillier Homomorphic encryption technique which can able to do the process in cipher text and generates result in scrambled form. It does the different computations in different services without revealing the data to each other. Only user is able to decrypt the result by using his own private key. Our IRIS dataset contains decimal values. As Paillier Homomorphic encryption scheme only supports integer, we scale decimal values by large positive integer and then perform encryption.

E. Algorithm

Algorithm 1: Backpropagation Training Algorithm

- Let i be the input layer , j be the hidden
- layer and k be the output layer.
- O_k is actual output and t_k is expected output
- Wij = weight between i and j.
- n be the number of inputs to node j.
- n=1;
- randomly initialize the weights;
- run the network forward with input and calculate the output O.
- For each output node compute the error function

$$\delta_{k=O_k(1-O_k)}(t_k - O_k)$$

- For each hidden node compute the error function

$$\delta_{j=o_j(1-o_j)}\delta_k w_{jk}$$

- update the weights in backward order starting from those of the output layer

$$\Delta W = -l_r \delta_l o_{l-1}$$
$$\Delta \theta = -l_r \delta_l$$

- Then apply,

$$\mathrm{W} = \mathrm{W} + \Delta\mathrm{W}$$
$$\theta = \theta + \Delta\theta$$

- End;

Algorithm 2: Paillier Homomorphic Encryption

1.Key Generation

- Pick two large prime numbers p and q.
- Compute n= pq
- Let λ be a carmichael function such that,

λ(n) = lcm(p-1)(q-1)

- Randomly pick gϵZ_{n2}^* such that

L(g^λ mod n^2) is invertible modulo n

- Where L(u)=u-1/n
- n and g are public
- p and q (or λ) are private.

2. Encryption

- For plaintext x and resulting ciphertext y, select a random
- r ϵ Zn^*
- e_K (x; r) = g^m r ^{n} mod n^2

3. Decryption

- d_K (y) =L{(y^λmod n^2)/L(g^λmod n^2)}mod n

F. Mathematical Model

Let S is the Whole System Consists:
S={I,P, O}
Where,
I=Input
P= Process
O=Output
I={t, PuKU}
P= {SC,Tr,Tt,CD,GR}
O = [[DEC]]
Users side parameters = {PuKU, PrKDU, t}
Server side Parameters = { [[t]]}
U={u1,u2,..un}
UD={t1,t2,t3,..tn}
U = EPuKU{UD}
{t1,t2,t3…tn}→{[[t1]],[[t2]]…[[tn]]}
P={SC, Tr, Tt}
U=send{[[UD]]} to server
Then server ,
SS={Tt,CD,GR}
Output = [[Dec]]
U=DPrKU{[[Dec]]}

G. Variables Used

Table: Memorization Parameters

Symbol	Meaning
UD	Users Data
U	User
E ,[[]]	Encryption
D	Decryption
PuKU	Public key of user
PrKU	Private key of user

CD	Classification of data
SC	Scaling
GR	Generation of result
u1,u2,..un	No of Users
t1,t2..tn	Users Data
Tr	Train the network
Tt	Testing
DEC	Decision

.

IV. RESULT AND DISCUSSION

In this framework by utilizing the artificial neural network we are going to accomplish the better speed of calculation and in future will likewise attempt to accomplish the accuracy as comparable with SVM. In this system we are using IRIS-plant dataset which having 3 classes, 4 attributes and 150 instances. Out of 150 instances we are using 120 instances for training purpose and remaining 30 for testing purpose.

Results

Table: Some samples of dataset

Sepal Length	Sepal Width	Petal length	Petal width	Class (Bit string representation)
5.20	3.40	1.40	0.20	Setosa (1 0 0)
4.90	3.60	1.40	0.10	Setosa (1 0 0)
5.60	2.70	4.20	1.30	Versicolor (0 1 0)
5.70	3.00	4.20	1.20	Versicolor (0 1 0)
)6.80	3.00	5.50	2.10	Verginica (0 0 1)
6.40	3.20	5.30	2.30	Verginica (0 0 1)

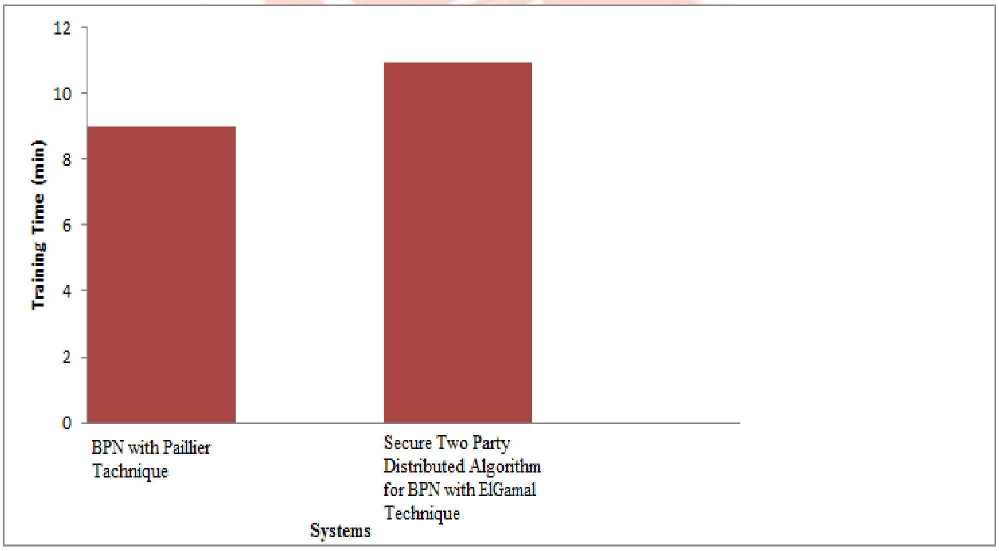

Fig. Training Time

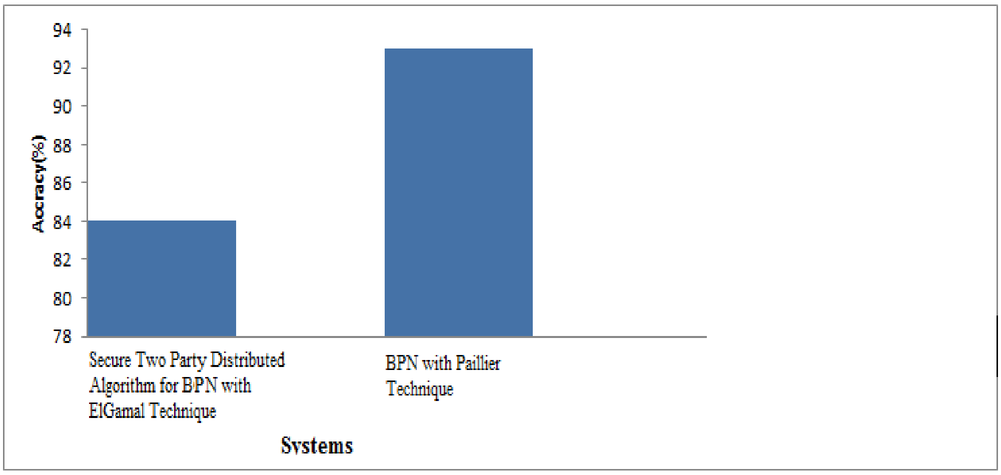

Fig. System Accuracy

V. CONCLUSION

The proposed system is a privacy preserving data classification system using artificial neural network which safeguards the security of client's information as well as jelly the server side choice and serve side choice making parameters. The system enhances the decision making capacity which can be successfully utilized as a part of any area. The system makes successful utilization of machine learning technique such as artificial neural network at server side for the purpose of classification at server side with the end goal of characterization which can bring about quick figuring speed.

VI. ACKNOWLEDGMENT

We are glad to express our sentiments of gratitude to all who rendered their valuable guidance to us. We would like to express our appreciation and thanks to the Principal of our college. We are also thankful to the Head of Department. We thank to the anonymous reviewers for their valuable comments.

VI. REFERENCES

[1]Yogachandran Rahulamathavan., "Privacy-Preserving Clinical Decision Support System Using Gaussian Kernel-Based Classification," IEEE journal of biomedical and health informatics,vol. 18, no. 1, january 2014.

[2] Pokharkar Shubhangi., "Review Techniques of Data Privacy in Cloud Using Back Propagation Neural Network", International Journal of Emerging Technology and Advanced Engineering ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 2, February 2014.

[3] Mr. Anwar Basha. "Privacy preserving back-propagation neural network learning using cloud computing," International Journal of Technical Research and Applications e- ISSN: 2320-8163, www.ijtra.com Volume 3, Issue 2 (Mar- Apr 2015), PP. 52-55.

[4] Arso M. Vukicevic , "Assessment of bone stress intensity factor using artificial neural networks ," IEEE conf., 2015

[5] Mrudula Gudadhe, Kapil Wankhade and Snehlata Dongre, "Decision Support System for heart disease using support vector machine and artificial neural network", IEEE conf., 2010.

[6] R. R. Janghel, Anupam Shukla and Ritu Tiwari, , "Clinical Decision Support System for Fetal Delivery using artificial neural network", IEEE conf ., 2009

[7] Sheng Zhong., "Privacy-Preserving Back propagation Neural Network Learning", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 20, NO. 10, OCTOBER 2009.

[8] P. J. Lisboa., "The use of artificial neural networks in decision support in cancer: A systematic review", Neural Netw., vol. 19,pp. 408415, 2006.

[9] M. Barakat., "Intelligible support vector machines for diagnosis of diabetes mellitus", IEEE Trans. Inf. Technol.Biomed., vol. 14, no. 4, pp. 11141120, Jul. 2010.

[10] N. Adhikari. " Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review", J. Amer. Med. Assoc., vol. 293,no. 10, pp. 12231238, 2005.

[11] E. R. Carson., "Clinical decision support, systems methodology, and telemedicine: Their role in the management of chronic disease", IEEE Trans. Inf. Technol. Biomed.,vol. 2, no. 2, pp. 8088, Jun. 1998.

[12] G. Mathew., "A privacy-preserving framework for distributed clinical decision support", in Proc. IEEE 1st Int. Conf. Comput. Adv. Bio. Med. Sci., 2011, pp. 129134.

[13] K.-P. Lin and M.-S. Chen, On the design and analysis of the privacy preserving SVM classifier, IEEE Trans.Knowl. Data Eng., vol. 23, no. 11, pp. 17041717, Nov 2011.

[14] D. Ubeyli, "Multiclass support vector machines for EEG signals classification", IEEE Trans. Inf. Technol. Biomed., vol. 11, no. 2, pp. 117126, Marb2007.

[15] M. Fung., "Anonymizing classification data for privacy preservation", IEEE Trans. Knowl. Data Eng., vol. 19, no. 5, pp. 711725, May 2007.

Ms. Sayali B. Desale currently pursuing a ME degree in computer science technology from Savitribai Phule Pune university. Completed B.E degree in computer engineering from R.H.Sapat college of engineering, Nashik under the Savitribai Phule Pune university, June 2014.

Prof. Santosh. B. Javheri received M. Tech in Information Technology from Bharati Vidyapeeth, Pune ,India in 2009. He Currently working as Associate Professor in Dept. of Computer Engineering, Rajarshi Shahu College of Engineering, Pune, India. He is having 17 years of teaching experience in various engineering college under SPPU Pune. His current research area is Data Mining and Information Security.