# K Nearest Neighbour Classification over Encrypted Relational Data

**Gadekar R.R.[1] Bhosale R.S.[2]**
[1]ME Student [2]Assistant Professor
[1,2]Department of Information Technology Engineering
[1,2]AVCOE, Sangamner, Maharashtra,India

*Abstract*— Data Mining widely used in many fields such as medical research, financial, medication and among govt. departments. For last years, query processing on relational data has been studied very much, and many solutions to query processing have been proposed under different scenarios. Classification is one of the applied works in data mining applications. As increasing use of cloud computing, users are able to outsource their information as well as the data management tasks to the cloud. This paper focus on different k-nearest neighbor (kNN) query problem over encrypted database outsourced to a cloud: a user issues an encrypted query record to the cloud and the cloud returns the k closest records to the user. This paper is motivated by need of secure classification technique to outsourced data to the cloud.

***Key words:*** Privacy, Outsourced Database, Encryption, KNN Classifier

## I. INTRODUCTION

As increasing popularity of cloud computing, users now able to outsource their information as well as management tasks to the cloud. Cloud computing has developed as a new platform for deploying, managing, and provisioning large-scale services through an Internet-based infrastructure such as Amazon EC2, Google App Engine. Last few years the cloud computing model is changing the organizations way of working their information. Almost all organizations assign their computational functions to the cloud to improve their information. Regardless of these advantages that the cloud environment offers, privacy and security issues in the cloud are thwarting companies to utilize those advantages. When data are private or highly sensitive, the data need to be encrypted before sending to the cloud. However, when data are encrypted, performing any data mining tasks becomes very challenging without ever decrypting the data. In this paper, survey various problem of secure processing of k-nearest neighbor query over encrypted data in the cloud.

### *A. Simple KNN:*

KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. To evaluate any technique we generally look at 3 important aspects:

1) Ease to interpret output
2) Calculation time
3) Predictive Power

Let us take a few examples to place KNN in the scale :

| | Logistic Regression | CART | Random Forest | KNN |
|---|---|---|---|---|
| 1. Ease to interpret output | 2 | 3 | 1 | 3 |
| 2. Calculation time | 3 | 2 | 1 | 3 |
| 3. Predictive Power | 2 | 2 | 3 | 2 |

Table 1:

KNN algorithm fairs across all parameters of considerations. It is commonly used for its easy of interpretation and low calculation time. KNN with distance algorithm is used in system.It uses Euclidean distance for calculation.

**Distance functions**

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

### *B. Paillier Cryptosystem:*

The Paillier cryptosystem is an additive homomorphic and probabilistic asymmetric encryption scheme.

Let Epk be the encryption function with public key pk given by (N; g),
where N is a product of two large primes and g.
Let Dsk be the decryption function with secret key sk.

Given a; b 2 ZN, the Paillier encryption scheme exhibits the following properties:

1) Homomorphic Addition
   Epk(a + b)← Epk(a) *Epk(b) mod N2;
2) Homomorphic Multiplication
   Epk(a *b) ←Epk(a)b mod N2;
3) Semantic Security - The encryption scheme is semantically secure, i.e., given a set of cipher texts, an adversary cannot deduce any information about the plain text.In this paper, we assume that a data owner encrypted his or her data using Paillier cryptosystem before outsourcing them to a cloud.

## II. IMPLEMENTATION

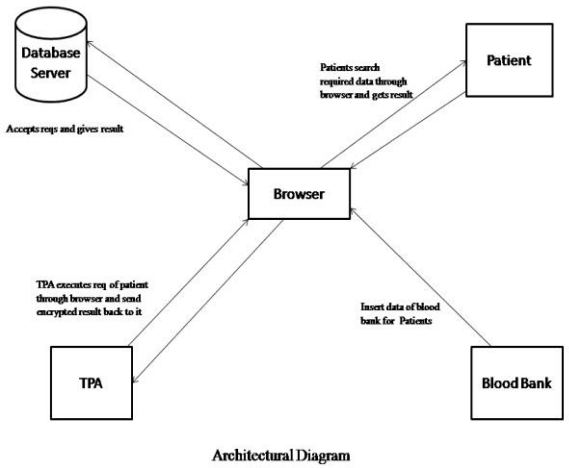


Fig. 1: Architecture of System

 

*A. Simple KNN:*

This algorithm is used in patient side in our system. It is the nearest neighbour algorithm. The *k*-nearest neighbour's algorithm is a technique for classifying objects based on the next training data in the feature space. It is among simplest of all mechanism learning algorithms. The algorithm operates on a set of d-dimensional vectors, $D = \{\mathbf{xi} \mid i = 1. . . N\}$, where $\mathbf{xi} \in kd$ denotes the *i* th data point. The algorithm is initialized by selection *k* points in *kd* as the initial *k* cluster representatives or "centroids". Techniques for select these primary seeds include sampling at random from the dataset, setting them as the solution of clustering a small subset of the data or perturbing the global mean of the data *k* times. Then the algorithm iterates between two steps till junction:

Step 1: Data Assignment each data point is assign to its adjoining centroid, with ties broken arbitrarily. This results in a partitioning of the data.

Step 2: Relocation of "means". Each group representative is relocating to the center (mean) of all data points assign to it. If the data points come with a possibility measure (Weights), then the relocation is to the expectations (weighted mean) of the data partitions.

Using this, Patient find out type of blood group by Rh value. patient give Rh+ value and by that type of blood group is find out.

*B. KNN with Distance:*

In this KNN used with Euclidean distance. this algorithm is used for find out nearest blood bank.

*C. Paillier Cryptosystem:*

This algorithm is used for encryption in this system.When patient request for data this algorithm internally encrypts data and send to the patient.then patient decrypts data.For this,it access tables also.

## III. Conclusions

To secure user privacy, numerous privacy-preserving category methods have been suggested over the past several years. The current methods are not possible to contracted database surroundings where the information exists in secured form on a third-party server. This paper suggested a novel privacy-preserving k-NN classification protocol over secured information in the cloud. Our protocol defends the privacy of the information, user's input query, and conceals the information access patterns. We also analyzed the efficiency of our protocol under various parameter configurations. Also, we will examine and increase our research to other category algorithms.

## References

[1] P. Mell and T. Grance, ―The NIST definition of cloud computing (draft), NIST Special Publication, vol. 800, p. 145, 2011.
[2] S. De Capitani di Vimercati, S. Foresti, and P. Samarati, ―Managing and accessing data in the cloud: Privacy risks and approaches, in Proc. 7th Int. Conf. Risk Security Internet Syst., 2012, pp. 1–9.
[3] P. Williams, R. Sion, and B. Carbunar, ―Building castles out of mud: Practical access pattern privacy and correctness on untrusted storage,in Proc. 15th ACM Conf. Compute. Common. Security, 2008, pp. 139–148.
[4] P. Williams, R. Sion, and B. Carbunar, ―Building castles out of mud: Practical access pattern privacy and correctness on Untrusted storage, in Proc. 15th ACM Conf. Compute. Common. Security, 2008, pp. 139–148.
[5] C. Gentry, ―Fully homomorphic encryption using ideal lattices, in Proc. 41st Annu. ACM Sympos. Theory Compute. 2009, pp. 169– 178.
[6] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, ―k-nearest neighbor classification over semantically secure encrypted relational data, e-print arXiv: 1403.5001, 2014.
[7] C. Gentry and S. Halevi, ―Implementing gentry's fully-homomorphic encryption scheme,in Proc. 30th Annu. Int. Conf. Theory Appl. Cryptographic Techno. Adv. Cryptol., 2011, pp. 129–148.
[8] A. Shamir, ―How to share a secret,Common. ACM, vol. 22, pp. 612–613, 1979.

 