IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
*Research Paper* **© 2012 IJFANS. All Rights Reserved**, Journal Volume 11, Iss 12, 2022

# K-NEAREST NEIGHBOR CLASSIFICATIONOVER SEMANTICALLY SECURE ENCRYPTEDRELATIONAL DATA

**[#1]Dr. Sk.Yakoob, Associate Professor,**
**[#2]Ch.Balakrishna, Assistant Professor,**
**[#3]J.Raja Kala, Assistant Professor,**
**Department of Computer Science and Engineering,**
**SAI SPURTHI INSTITUTE OF TECHNOLOGY, SATHUPALLY, KHAMMAM.**

## ABSTRACT:

Data mining is used in many contexts, such as business, healthcare, science, and government. Data mining frequently makes use of categorization. Recent years have seen a meteoric growth in the development of both theoretical and practical classification systems in response to rising privacy concerns. These changes were implemented using a wide variety of security frameworks. As cloud computing grows in popularity, more and more people are turning to it for secure data storage and transmission to remote servers for data mining. Data encryption in the cloud renders obsolete the existing privacy protection classification methods. The purpose of this study is to identify a solution to the issue of encrypted data classification. For encrypted cloud-stored data, we recommend utilizing a powerful k-NN classifier. The suggested protocol is designed to protect the confidentiality of users' search histories and access logs. To our knowledge, this is the first research to employ the semi-honest model in order to develop a trustworthy k-NN classifier capable of processing encrypted data. Furthermore, we use an observational dataset with multiple adjustment factors to evaluate our proposed approach in the actual world.

Encryption, k-NN classifier, cloud-based data storage, and safety are some of the terms employed.

***Keyword:*** **-** Security, k-NN classifier, outsourced databases, encryption

## 1. INTRODUCTION

Cloud computing has revolutionized how businesses store, access, and process data, rendering obsolete previous approaches. Flexible, inexpensive, and requiring less administration labor, cloud-based processing is gaining popularity among small and medium-sized organizations. More and more companies are realizing that by connecting their gadgets to the cloud, they can improve the reliability of their data. Many organizations avoid cloud computing due to worries about data security and user friendliness. It's best practice to encrypt sensitive information before storing it on the cloud. No matter how stringently data is protected, gaining access to encrypted data without cracking the encryption is extremely challenging. This occurs because it is impossible to decipher the encrypted data. If you needed any more convincing to keep your distance, think about what happened next.

Consider an insurance firm that has outsourced the task of extracting and securing client data to a cloud service provider. A potential client's risk level can be assessed by a company rep utilizing a classification structure. The primary responsibility of the representative is to create a questionnaire to collect sensitive client data. Fill out this form with the consumer's personal information, including credit history, age, and marital status. The historical information can then be saved in the cloud and used to determine the q class label for the following observation.

Since the variable q may contain sensitive information, it must be encoded before being uploaded to the cloud. Only if you take this

 

precaution can you ensure the privacy of your customers' sensitive data.

Protecting a user's browsing history is crucial for DMED, or data mining on a cloud platform using encrypted data. The cloud, despite its security safeguards, can nonetheless learn private information about some things by monitoring their use. Encrypting data, protecting user search history, and concealing how people access information are three key privacy and security needs that must be met in order to solve DMED in the cloud.

Neither perturbation-based methods nor an emphasis on secured multi-party computation (SMC) are doing enough at the moment to solve the DMED challenge in privacy-preserving data mining (PPDM). Due to a lack of semantic safeguards, information perturbation approaches are not suitable for extremely sensitive content, hence the name "Protecting Personal Data Mining." Inaccurate information throughout the system makes research more difficult. Since private information must be shared across many parties, security cannot be ensured for all participants in a secure multiparty computation. Numerous extremely intricate computations rely on data that has not been encrypted. According to this paper, encrypted information is stored in the cloud. We then propose a wide variety of novel approaches to the DMED issue. The category problem is crucial since it is one of the most widely used data mining methods. The k-nearest neighbor classification algorithm is used to read encrypted files in a cloud computing environment. The major guidelines and advantages of several classification systems are outlined in this article.

## 2.LITERATURE SURVEY

Data can be stored in a secure and convenient off-site location, as demonstrated by this study. This strategy simplifies data collection, is consistently effective, and excels when applied to familiar patterns. Data read, written, and added by a storage client using this method remains their own and is protected from prying eyes. It is effective even if you are working with a storage service provider who is either avaricious or malicious. There is no way for the service provider to distinguish between reading and writing, or to link together multiple successive accesses. In addition, buyers are assured of the items' quality and functionality, and any unethical or illegal business activities are prohibited. We've developed state-of-the-art technology that can efficiently process thousands of queries per second on databases larger than one terabyte. When compared to other products, this one stands head and shoulders above the rest.

This research demonstrates that circuits may be verified with encrypted data without knowing how to decrypt it by employing a completely homomorphic security technique. There are three phases to the response, and they must be completed in the specified order. Whatever security mechanism is used to identify unnecessary circuits must also be capable of verifying the operation of its own decryption circuit, no matter how complex it may be. Over the next three weeks, we'll examine an extremely self-sufficient method of protecting public keys using perfect lattices. Lattice-based cryptosystems typically employ straightforward decoding techniques because of the usage of perfect lattices, making this approach "bootstrappable." A class NC1 inner item computation is used in several of these techniques. Lattices representing polynomial bands have preservative and multiplicative homeomorphisms if and only if the public key is perfect. Traditional circuit testing relies on the existence of these homeomorphisms, which are unique to perfect lattices. Following the instructions in this article, you can easily split D into n pieces and reassemble it from any of the k pieces. It is also essential to remember that D cannot be inferred from the characteristics of the

 

k-1 components. This method simplifies the development of safe and efficient cryptographic key management operations. These techniques are still effective even after fifty percent of the items have been misplaced or stolen and all but one have been hacked.

When it comes to collecting and monitoring sensitive personal data, paper documents present a significant security risk. It is not only impossible, but also completely unreasonable, to build a universally accessible computer using a single formula. This paper provides an efficient mathematical approach to solving the issue at hand. Sharemind is an online PC where users may pool their resources secretly and securely.

Methods like these are used to locate vulnerabilities in features that allow numerous users to use a single computer. Our solution is novel because of the deliberate choices we took when disseminating sensitive data and assembling the protocol bundle. There are a number of effective methods available to assist students cope with the heated debates that frequently arise during training. According to the SHAREMIND protocol, all three executives are reliable and totally committed to their jobs. The design of the honest-but-curious database makes it more fun to use than traditional centralized databases while still keeping malicious users out.

This essay focuses on methods for data mining that respect people's right to privacy. In this scenario, two companies, each with its own private database, would like to utilize a data mining technique on a combined dataset without disclosing any of the data in either database. Knowing that their data will be used for research and other purposes motivates people to give their all. Using a standard and comprehensive approach to encrypted multi-party computations, you can solve your challenge. Due of the copious amounts of information they generate, data mining techniques can be challenging to discover and implement. As a result, we need to abandon the tried-and-true approach to problem resolution in favor of more novel strategies. Our primary goal is to collect a large amount of information regarding decision trees using the widely used ID3 technique. Our system outperforms competing approaches while requiring fewer interaction units and less bandwidth for data delivery. Therefore, it is a great choice.

This research investigates the best practices for discovering product linkages by analyzing historical data on individual product sales. The secrecy of monetary dealings is safeguarded by the results derived from randomly generated data. Regrettably, the same public regulations that can be utilized to locate rule breakers can also be used to punish them. While a simple randomization method might help an organization uphold its ideals and speed up its procedures, it's unfortunate that rule infractions can be uncovered in this way. Find out what kinds of privacy invasions are possible, and then come up with a randomization technique that is considerably better at preventing those kinds of invasions than regular randomization is. The neutral support estimator's variance and formula are found afterward. This allows us to demonstrate the utility of these equations for analyzing methods and facilitating item sets using simulated data. Experiments employing the criterion revealed positive outcomes when applied to real-world data.

Since databases efficiently organize data and simplify human interaction with systems, they speed up the process of problem fixing with convenience. Using data warehousing and data mining to compile information from various sources can help reduce the likelihood of a comfort crime. Data mining techniques that safeguard individual privacy by revealing only the desired result may provide a solution to these issues. This article explains how to work with k-nearest neighbor (k-nn) classes manually,

without the need for any specialized software or hardware. The approach promises that only the final category decision, and not any information concerning assigned resources or related data, will be disclosed.

The study's findings highlight the need for data mining techniques that safeguard consumers' confidentiality when compiling information from various databases. To safely and securely explore nearby partitioned databases without compromising users' privacy, a comprehensive framework, a detailed model, and iterative approaches based on the k Nearest Neighbor (kNN) classifier are described. Modeling, algorithms, and iterative algorithms are the three pillars of the framework.

The study found that encrypted data might be used to perform searches for multidimensional variety. The issue arises because so many individuals rely on tools that allow them upload sensitive information to a distant server and then utilize that data and the server's resources for something else. Freelancers who value their privacy and independence utilize these apps. The proposed fix employs a data-splitting technique called bucketization to accurately label the information. The server has no access to the values but can determine if the record satisfies the criteria specified. Incomplete questions can lead to incorrect data being returned. We shift the onus of our solution's computation to the user by eliminating unnecessary information. The purpose of this research is to develop a bucketization technique that can simplify the processing of multidimensional data. A unique method of bucketization is used to provide analytics for price and disclosure risk. These figures also reveal the time and money invested in client computing, as well as the potential for sensitive information to fall into the wrong hands. When dealing with diverse information, bucketization is a challenge in the marketing industry. Protecting sensitive data while limiting the amount of effort required on the client's PC is the goal. Data owners can weigh the benefits and drawbacks of sharing their data with others using the several categorization choices provided by our service. We also run several experiments with simulated and actual data to better understand the tradeoffs. The SaaS business is becoming increasingly competitive, despite the fact that Google and Amazon are making significant strides forward. The company has converted its massive data centers into cloud computing environments and is actively courting companies interested in using its servers to host their applications. For this service model to be effective and secure, it is crucial that all data processed within the system be kept private at all times. Standard security measures, despite their best efforts, do not improve the efficiency of applications that employ protected data, such as database searches. This research demonstrates the difficulty of securely calculating using encrypted databases. SCONEDB, which stands for Secure Computation ON an Encrypted Database, is recommended by experts as a method that effectively combines speed with security. The primary objective is to identify the k-nearest neighbors (kNN) of a secret dataset. This study demonstrates the use of ASPE, an asymmetric scalar-product-preserving encryption algorithm. This substance safeguards a crucial category of scalar objects against harm.

We provide two secure kNN computation methods for APSE-encrypted data. Both approaches have proven effective at preventing attacks by adversaries with varying levels of expertise and equipment, while they may require distinct investments of time and money. The procedures' efficiency and utility are rigorously examined.

## 3.CONCLUSIONS

Data mining has many applications, but its primary goal is data classification, such as detecting credit card fraud or counting cancer cells

in a patient's body. Many innovative privacy-protecting classification systems have emerged in the last decade. We need better tools to deal with encrypted data in the cloud while using a database service. This research demonstrates how to use encrypted cloud data for secure k-NN classification. Our protocol ensures the privacy of all sent information, including user inquiries and data. Different combinations of parameters were used to evaluate our method's performance.

## 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. Shamir, ―How to share a secret,‖ Common. ACM, vol. 22, pp. 612–613, 1979.
8. D. Bogdanov, S. Laur, and J. Williamson, ―Sharemind: A framework for fast privacy-preserving computations,‖ in Proc. 13th Eur. Symp. Res. Compute. Security: Compute. Security, 2008, pp. 192–206.
9. Y. Lindell and B. Pinkas, ―Privacy preserving data mining,‖ in Proc. 20th Annu. Int. Cryptol. Conf. Adv. Cryptol., 2000, pp. 36–54.
10. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, ―Privacy preserving mining of association rules,‖ Inf. Syst., vol. 29, no. 4, pp. 343–364, 2004.