Comparative Study of Knn/Pcs with Naive Bayes/Rsa

Dr.Savithri.V & Ms.Aruna
Asst. Professor & Research Scholar
Women's Chrisitian College & M.T.W.U., Kodaikanal

Abstract: Data mining, the extraction of concealed predictive data from substantial databases, is an intense new innovation with awesome potential to help organizations concentrate on the most important data in their data warehouses. Data processing tools predict future trends and behaviors, permitting businesses to create proactive, knowledge-driven choices. During this paper, managing the classification issue over disorganized data. Specifically, during this planned system protected Naïve Bayes classifier is employed to classify the info set and RSA cryptosystem is employed to cipher and decode the info set within the cloud. RSA may be a cryptosystem for public-key encoding, and is wide used for securing sensitive knowledge, significantly once being sent over an insecure network like the web. The planned convention ensures the confidentiality of data, security of client's data inquiry, and conceals the data access styles. To the most effective of our insight, this work is that the first to creat up a secure naïve bayes classifier over encoded data beneath the semi-fair model.

Keywords: Naive Bayes classification, RSA cryptosystem, encryption.

1. INTRODUCTION

Data mining is that the method of discovering purposeful new correlations, patterns and trends separation through massive amounts of knowledge hold on in repositories, victimization pattern recognition technologies further as applied statistical and mathematical techniques. Data mining is associate repetitive and interactive process of discovering one thing innovative.

Data mining consists of a collection of techniques that may be used to extract relevant and fascinating knowledge from data. It has many tasks like association rule mining, classification and prediction, and clustering. Classification techniques are supervised learning techniques that category data item into predefined class label. It's one among the foremost helpful techniques in data processing to create classification models from associate computer file set. The used classification techniques commonly build models that are used to predict future data trends. There are many algorithms for data classification like decision tree and naïve bayes classifiers. With classification, the generated model are going to be ready to predict a category for given data depending on previously learned information from historical data.[10]

Classification is that the most often used data processing task with a majority of the implementation of Bayesian classifiers, neural networks, and SVMs (Support Vector Machines). A myriad of quantitative performance measures were projected with a predominance of accuracy, sensitivity, specificity, and ROC curves. Naïve Bayes classifier is another classification technique that's utilized to predict a target class. It depends in its calculations on possibilities, particularly Bayesian theorem. Due to this

use, results from this classifier more exact, effective, and additional sensitive to new knowledge added to the dataset. There are three dimensions that we must consider with a privacy preserving classifications formula, namely, accuracy, efficiency, and privacy. [7] Privacy-perserving data processing becomes a very important enabling technology for mining knowledge from multiple private databases provided by different and probably competitory organizations.

Encryption is a settled innovation for ensuring delicate information. Unfortunately, the incorporation of existing encryption methods with database frameworks causes undesirable execution corruption. For instance, if a section of a table containing delicate data is encoded, and is utilized as a part of an inquiry predicate with an examination administrator, a whole table output would be expected to assess the question. The reason is that the present encryption strategies don't secure request and accordingly database records, for example, B-tree can never again be utilized. In this way the question execution over encoded databases can turn out to be unsuitably moderate.[8] Completely homomorphic encryption plan is a public key encryption plan utilizing perfect lattices that is practically boots trappable. lattice based cryptosystems commonly have decoding calculations with low circuit complexity, regularly dominated by an inner item calculation that is in NC1. Additionally, ideal lattices give both added substance and multiplicative homomorphisms (modulo an open key perfect in a polynomial ring that is pictured as lattice), as expected to assess general circuits. [6]. Since the SVD (secure voronoi outline) strategy does not utilize any new encryption plans, rather, it just depends on any standard encryption plan E (e.g., public key encryption RSA, symmetric-key encryption AES), the SVD technique is as secure as E for any standard security model in which E is demonstrated secure (e.g., indistinctness in either picked plaintext or picked cipher text assaults). [10]

2. RELATED WORKS

[1] Making sure right privatives and safety of the statistics saved, communicated, processed, and disseminated within the cloud in addition to of the customers having access to such an data is one of the grand challenges of our cutting-edge society. As a rely of fact, the improvements in the statistics generation and the diffusion of novel paradigms such as facts outsourcing and cloud computing, whilst allowing customers and businesses to easily get entry to high excellent programs and services, introduce novel privacy dangers of unsuitable records disclosure and dissemination. In this paper, we will symbolize special aspects of the privacy hassle in emerging situations. This research will illustrate risks, answers, and open problems associated with making sure privatives of customers gaining access to offerings or assets in the cloud, sensitive information stored at external events, and accesses to such a statistics.

[2]This paper presents another useful component for remote information stockpiling with efficient access design protection and accuracy. A capacity

customer can send this system to issue scrambled peruses, composes, and embeds to a possibly inquisitive and malevolent capacity administration supplier, without uncovering data or access designs. The supplier can't build up any connection between progressive gets to, or even to recognize a read and a compose. Also, the customer is given with solid accuracy confirmations to its operations unlawful supplier conduct does not go undetected. This research manufactured a first handy framework request of greatness quicker than existing usage that can execute more than a few questions for each second on 1Tbyte+ databases with full computational security and accuracy.

[3]This paper explores a novel computational issue, to be specific the Composite Residuosity Class Problem, and its applications to open key cryptography. This research propose another trapdoor instrument and get from these method three encryption plans: a trapdoor stage and two homomorphic probabilistic encryption conspires computationally equivalent to RSA. Our cryptosystems, in view of regular particular mathematics, are provably secure under proper presumptions in the standard model.

[4]The prerequisite for open key cryptosystem is spreading quickly today when more individuals use PC systems to safely trade mystery data. Because of the significance of open key cryptosystem, there are numerous the general population key plans which are fundamentally taking into account the hardness suspicions, for example, considering, discrete logarithm or different cross section issues, individually. To date, all cryptographic plans just are endless supply of three sort issues. At present, there does not exist the cryptographic primitive all the while in light of the hardness presumptions of the issues from various territories.

[5] Data Mining has wide applications in numerous zones, for example, banking, medication, scientific research and among government offices. Classification is one of the ordinarily utilized task as a part of information mining applications. For as far back as decade, because of the ascent of different protection issues, numerous hypothetical and down to earth answers for the classification issue have been proposed under various security models. In any case, with the late prominence of distributed computing, clients now have the chance to outsource their information, in scrambled structure, and in addition the information mining errands to the cloud. Since the information on the cloud is in encoded structure, existing protection saving classification methods are not relevant. In this paper, concentrate on tackling the classification issue over encoded information. Specifically, this research propose a safe K-NN classifier over scrambled information in the cloud. The proposed convention ensures the confidentiality of information, protection of client's info inquiry, and conceals the information access designs. To the best of our insight, our work is the first to build up a safe k-NN classifier over scrambled information under the semi-legit model. Additionally, we observationally investigate the efficiency of our proposed convention

utilizing a certifiable dataset under various parameter settings.

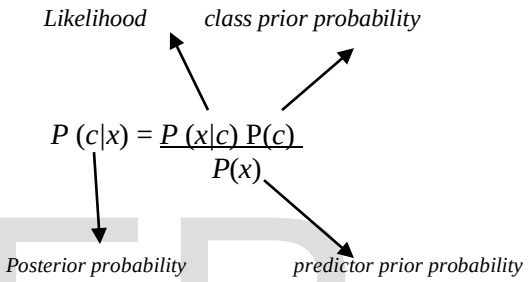
3. METHODOLOGY

Classification is one in every of the foremost vital variable techniques employed in statistics. It's closely associated with prediction and interestingly the classification drawback is typically known as the prediction issue significantly in data processing. There are many approaches those deals with the classification issue. The statistical based algorithm Naïve Bayes Classifier is often employed in prediction issue. Just in case of Naïve Bayes formula one in every of the most factors is to influence numerical attributes. It's obvious as a result of within the formula one should verify the conditional probability for every possible value of all attributes. To resolve this issues, we've got to discretize numerical attributes into many classes by adopting the discretization technique from a good vary of choices assessable. Therefore the techniques used for discretization plays a very important role over the accuracy of the strategy. Several attempt are created to extend the accuracy of the Naïve Bayes algorithm by adopting a brand new discretization theme.

The Naïve Bayesian classifier relies on Bayes theorem with independence assumptions between the predictors. A Naïve Bayesian model is simple to make, with no sophisticated iterative parameter estimation that makes it especially helpful for massive information sets. Despite its simplicity, the Naïve Bayesian classifier typically does surprisingly well and is wide used as a result of it typically outperforms a lot of sophisticated classification methods.

3.1 Algorithm

Bayes theorem provides the simplest way of calculating the posterior probability, P(c/x) from P(c), P(x), and P(x/c). Naïve Bayes classifier assumes that the impact of the worth of a predictor (x) on a given class (c) is independent of the values of different predictors. This assumption is named conditional independences.

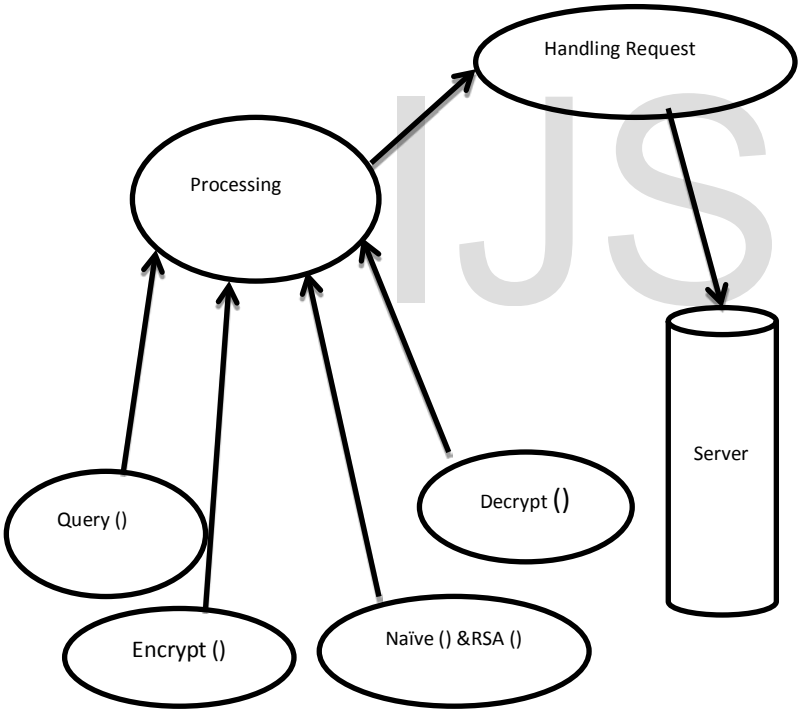


$$P(c \mid x) = P(x_1 \mid c) \times P(x_2 \mid c) \times \ldots \times P(x_n \mid c) \times P(c)$$

P(c | x) is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
P (c) is the prior probability of *class*.
P(x | c) is the likelihood which is the probability of *predictor* given *class*.
P(x) is the prior probability of *predictor*.

RSA was initially depicted in 1977 by Ron Rivest, Adi Shamir and Leonard Adleman of the Massachusetts Institute of Technology. Open key (public key) cryptography, otherwise called asymmetric cryptography, utilizes two distinctive however numerically connected keys, one open and another private. General public key can be common to everybody, though the private key must be kept mystery. In RSA cryptography, both open and the

private keys can encode a message; the inverse key from the one used to encode a message is utilized to decode it. This quality is one motivation behind why RSA has turned into the most generally utilized asymmetric calculation. It gives a strategy for guaranteeing the classification, integrity, authenticity and non-reputability of electronic interchanges and information storage. RSA executes an open key cryptosystem that permits secure communications and its security lays to some extent on the trouble of calculating huge numbers.



3.2 RSA implemented two important ideas:

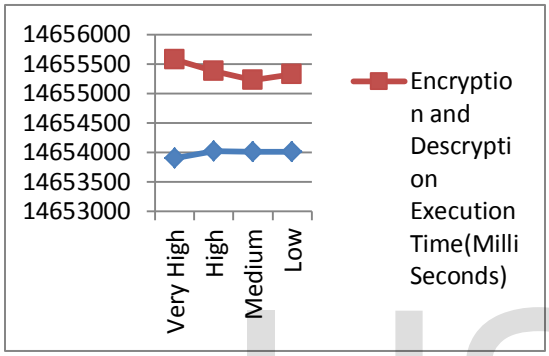
1. Public-key Encoding. This thought omits the requirement for a "dispatch" to deliver keys to recipients over another safe channel before transmitting the initially originally-planned message. In RSA, encoding keys are public, while the decoding keys are not, so just the individual with the right decoding key can decipher an encoded message. Everybody has their own particular encoding and decoding keys. The keys should be created in a manner that the decoding key might not be simply derived from the public encoding key.

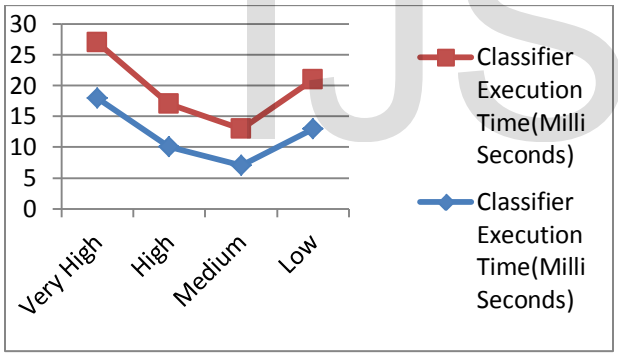
2. Digital signatures. The receiver may have to check that a transmitted message truly started from the sender (signature), and didn't simply come back from there (confirmation). This is can be done utilizing the sender's decoding key, and also the signature will later be verified by anyone, utilizing the corresponding public encoding key. Signature hence can't be produced. Also, no signer will later deny having signed the message. This is helpful of electronic mail as well as for other transactions and transmissions, like money transfers. The protection of the RSA calculation has so far been approved, since no known makes an attempt to interrupt it have yet been successful, for the most part because of the problem of factoring huge numbers n= pq, where p and q are large prime numbers.

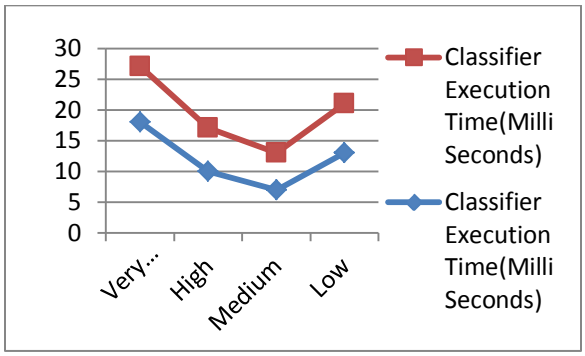
4. CONCLUSION.

To protect user privacy, numerous privacy-preserving classification techniques have been planned over the past decade. The existing techniques are not applicable to outsourced database environments where the data resides in encrypted form on a third-party server. This paper proposes a novel privacy-preserving classification protocol over encrypted data in the cloud.

Car cost	Encryption and Decryption Execution Time (Milli Seconds) Existing	Encryption and Decryption Execution Time (Milli seconds) Proposed
Very High	14653908	1665
High	14654019	1360
Medium	14654012	1218
low	14654008	1314







In this paper proposes a Naive Bayes classification with Rsa cryptosystem as an alternative to Knn classification with Paillier cryptosystem. Experiments on a large number of datasets show that the proposed system has given the result with more accuracy, effective and encryption time consumes very less compared to existing system. We plan to investigate alternative and more efficient solutions to the Security problem in our future work. Also, we will investigate and extend our research to other classification algorithms.

REFERENCES

[1] 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.

[2] 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. Comput. Commun. Security, 2008, pp. 139–148.

[3] P. Paillier, "Public key cryptosystems based on composite degree residuosity classes," in Proc. 17th Int. Conf. Theory Appl. Cryptographic Techn., 1999, pp. 223–238.

[4] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "k-nearest neighbor classification over semantically secure encrypted relational data," eprint arXiv:1403.5001, 2014.

[5] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in Proc. 30th Annu. Int. Conf. Theory Appl. Cryptographic Techn.: Adv. Cryptol., 2011, pp. 129–148.

[6] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. 41st Annu. ACM Sympos. Theory Comput., 2009, pp. 169– 178.

[7] L. Xiong, S. Chitti, and L. Liu, "K nearest neighbor classification across multiple private databases," in Proc. 15th ACM Int. Conf. Inform. Knowl. Manage., 2006, pp. 840–841.

[8] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2004, pp. 563–574.

[9] Qasem A. Al-Radaideh, Eman Al Nagi , "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance". International Journal of Advanced Computer Science & Application;Feb2012, Vol. 3 Issue 2, p144

[10] X. Xiao, F. Li, and B. Yao, "Secure nearest neighbor revisited," in Proc. IEEE Int. Conf. Data Eng., 2013, pp. 733–744.