Secure Cloud Model using Classification and Cryptography

Tamanna
M.Tech Scholar
CSE Department
DAV Institute Of
Engineering and Technology
Jalandhar,India

Rajeev Kumar
Assistant Professor
IT Department
DAV Institute Of
Engineering and Technology
Jalandhar, India

ABSTRACT
Cloud computing offers numerous benefits including scalability, availability and many services. But with its wide acceptance all over the globe, new risks and vulnerabilities have appeared too. Cloud computing supplies facility of storing and accessing understanding and programs over the web without bothering the storage space on procedure. Storing the data on cloud eliminates one's worries about space considerations, buying new storage equipment or managing their data, rather they are able to access their data any time from any place provided they have internet access. However, the rising security issues have resisted the companies from connecting with cloud computing fully. Hence security risks have appeared as the main disadvantage of cloud computing. This paper involves the efforts to research the security risk and then proposes a framework to address these risks on the authentication and storage level in cloud computing. While addressing the security issues the first and the foremost thing is to classify what data needs security and what data needn't bother with security and hence data gets classified into classes. To achieve data classification, a data classification approach based on the confidentiality of data is proposed in this paper. Following that an efficient security mechanism must be deployed by means of encryption, authentication, and authorization or by means of every other method to ensure the privacy of s data on cloud storage

Keywords
Confidentiality, Privacy Preserving, Machine learning, data classification, KNN and Naïve Bayes

1. INTRODUCTION
In at current technology of competitors, businesses are below huge pressure to reinforce effectivity and transform their IT procedures to gain extra with much less. Businesses need decreased time-to-market, higher availability, higher agility, and lowered bills to fulfill the difficult industry specifications. All these challenges are addressed through new computing style known as cloud computing.

Cloud Computing is an internet founded allotted digital atmosphere. All computational operations are carried out on cloud through the internet. The rate of the resource administration is greater than the specific fee of the assets. So, it is often better to get the desired resources by way of renting as a substitute of purchasing one's possess resources.

Essentially, the cloud computing presents all IT assets for hire. The definition of cloud computing is: "A distributed virtual environment provides virtualization based IT-as-Services by rent". Beside all of the services like application-as-a-service (SaaS), Platform-as-a-carrier (PaaS) and Infrastructure-as-a- carrier (IaaS), cloud also presents storage as a provider, where dispensed database servers are on hand for rent to shoppers [1]. These services are to be had for all users without any knowledge bias. With cloud computing, users can browse and decide upon critical cloud offerings, such as laptop, software, storage, or combo of these resources, by way of a portal. Cloud computing automates supply of selected cloud offerings to the customers. It helps the corporations and contributors installation IT resources at decreased whole cost of possession with rapid provisioning. As cloud computing helps companies to sharpen their development and performance. Besides this, it also hosts many users to furnish entry to shared resources with less effort. But security issues or threats are nonetheless a stumbling block in the success route of cloud computing. Numbers of factors are the subject. First intent is that users and many businesses store their knowledge on cloud storage, so the most important focus is the info ought to be comfy, and the info are not being lost and tampered even as traveling from one situation to yet another over the network. So it is main that confidentiality, availability and integrity of data will have to be ensured. Secondly, unauthorized access where an attacker tries to be the impersonator of the legal person. [2]

Security is the number one limitation in relation to any upcoming science and cloud computing is not any exception. Cloud computing poses countless security risks in allotted cache mannequin. Know-how protection is the major setting up hazard for the nature of administrations that forestalls the consumers to embrace the cloud administrations. In disbursed storage, the knowledge is put away on the separates by way of two cache systems. The previous is to encode the knowledge and store on the server at the same time the last is to store the know-how without encryption. These functions can often face confidentiality issue. The data is regularly not of the same sort and may have distinctive properties. As the consumer's data is stored on the remote servers and the consumer has no idea about its physical location, so there is always a risk of confidentiality leakage. [3] This paper concentrates on privacy dilemma in cloud computing. At whatever factor the know-how is exchanged to the cloud server it experiences a security system i.e. Encryption without comprehension the extent of sensitivity of the information or the info is basically put away on cloud server without securing it. All understanding has numerous sensitivity phases so it is incorrect to store the know-how without comprehension its sensitivity stage and protection requirements. To direct the security necessities of data, now we have proposed an information classification mannequin to classify the data according to its sensitivity stage and then encrypting the one knowledge which is

required to comfy using an encryption procedure in cloud atmosphere.

Classification of objects is an imperative field of research and of practical applications in numerous fields like pattern recognition and artificial intelligence, statistics, vision analysis and remedy. . A very intelligent technique to secure the data would be to first classify the data into sensitive and non-sensitive data and then secure the sensitive data only. This will help to reduce the overhead in encrypting the entire data which will be exceptionally costly in connection of both time and memory. For encrypting the data many encryption techniques can be used and for classifying the data numerous classification algorithms are to be had in the discipline of data mining.

Data classification is a laptop studying process used to foretell the category of the unclassified understanding. Knowledge mining makes use of specified instruments to grasp the unknown, respectable patterns and relationships within the dataset. These tools are numerical calculations, factual models and prediction and evaluation of the info. Hence, data mining contains management, collection, prediction and analysis of the data. ML algorithms are described in to 2 classes: supervised and unsupervised. In supervised studying, courses are already outlined. For supervised studying, first, an test dataset is defined which belongs to individual courses. These lessons are appropriately labelled with a specific identify. Lots of the data mining algorithms are supervised finding out with a designated intention variable. In unsupervised learning classes aren't without difficulty characterized but as a substitute arrangement of the know-how is performed automatically. The unsupervised algorithm looks for similarity between two gadgets in order to find whether they are able to be characterized as forming a cluster. In simple words, in unsupervised learning, "no goal variable is identified". The classification of know-how within the context of confidentiality is the classification of expertise headquartered on its sensitivity level and the have an affect on to the organization that capabilities be disclosed handiest licensed users. The info classification helps investigate what baseline security standards/controls are correct for safeguarding that knowledge. The knowledge is labeled into two classes, personal and non-exclusive (non-distinct) understanding. The classification of the information relies on the attributes of the knowledge. The values of the sensitive attributes are labeled as "confidential"and "highly confidential " and values of the non-touchy attributes are categorized as "basic".

The remainder of this paper is organized as follows: In section 2, related work is mentioned. In section 3, proposed work is presented. In section 4, results and discussions are discussed. The document has been concluded in section 5 with future research directions.

2. RELATED WORK

Somani U et.al [4] In this RSA algorithm is used to ensure the confidentiality aspect of security whereas Digital signatures were used to enhance more security by authenticating it through Digital Signatures. The approach used carryout encryption in 5 steps. In first step, key is generated. In second step, digital signing is performed and in step 3 and step 4 encryption and decryption is carried out. In last step Signature verification is performed. **Rewagad P et.al [7]** They have proposed an architecture to protect confidentiality of data stored in cloud by making use of digital signature and Diffie Hellman key exchange with (AES) Advanced Encryption Standard encryption algorithm .Even if the key in transmission is hacked, the facility of Diffie Hellman key exchange make it useless because key in transit is of no use without user's private key, which is provided only to the legitimate user. This three way mechanism proposed architecture makes it tough for hackers to break the security system, thereby protecting data stored in cloud. **Sinha N et.al [8]** This paper provides the brief history of cloud computing with its benefits architecture implementation and all issues in cloud computing. It provides the basic idea of all the different kind of issues related to security, data and performance in cloud.

Diwan V et.al [9] This paper different cryptographic algorithm are been compared which are been taken into consideration to provide the confidentiality of the data.In this different cryptographic algorithms are being compared by considering different parameters like block size, key length type and features.This paper had provided the idea of different cryptographic algorithm which can be used to ensure the security of data in cloud. **Zardari MA et.al[10]** In this paper they had used the K-NN approach for in order to do classification of data inorder to provide the confidentiality of data.The main aim to classify data is to provide security.In this approach they classify the data into labels sensitive and non-sensitive data using K-NN algorithm.On the sensitive data the encryption is done inorder to provide the security.Classification is mainly performed because it become easy to select an appropriate security for data according to need of data.So this way it will enhance the security. **Shaikh, Rizwana et.al [11]** This paper had contributed another important technique inorder to enhance the security of data in cloud by considering the classification technique.In this different parameters are been taken into consideration to provide the classification of data and then on the classified data the encryption is performed.Different classification properties considered are access control, content and storage.On these properties the data classification is done and then the encryption is performed to enhance more security and better efficiency.**Tawalbeh L et al. [12]** In this paper they had contributed the secure cloud computing model based on the classification which basically minimizes the overhead and processing time needed to secure the data through using different security mechanism with variable keysizes to provide the confidentiality level required for data.Classification is done by user manually and the encryption is been performed in three different level with different cryptographic algorithm.The levels are based on the sensitivity of data which includes Basic, Confidential and highly confidential level.The different cryptographic algorithm are been used at different levels to provide the security of data upto a great extent.

3. PROPOSED WORK

The research involves exploring various data classification algorithms in machine learning like KNN, Naïve Bayes and improved Naïve Bayes algorithm and analyzes their performance.

The paper proposes a secure data classification model using novel Bayesian supervised machine learning approach. In this, data is classified according to its sensitivity level. Then encrypting only, the data which is required to be secure using a cryptographic technique in cloud environment.

3.1 Data Classification

Apply the improved Naïve Bayes algorithm for classification.

1. Combining naïve bayes with Decision table using Decision tree as Meta classifier.

2. Meta Learner is a learner scheme that combines the output of the naive bayes and decision table i.e. the base learner. The base learners' level-0 models and the meta-learner is a level-1 model. The predictions of the base learners are input to the meta-learner.

3.2 Data Encryption Architecture

Basic: Uploading data on cloud securely by encrypting it using AES algorithm.

Confidential: Uploading data on cloud securely by encrypting its data by using RSA algorithm

Highly confidential: Uploading data on cloud securely by encrypting its data using combined elgammal and hashing algorithm.

The key objective of the proposed algorithm is to obtain better results than the existing KNN machine learning algorithm on the basis of classification time, accuracy and also enhancing the security at confidentiality and integrity levels of cloud.

The evaluation parameters considered for evaluating the performance of the proposed system are:

a) Time taken for classifying the data

b) Accuracy of the classified data

c) True Positive rate

d) Encryption Time

e) Decryption Time

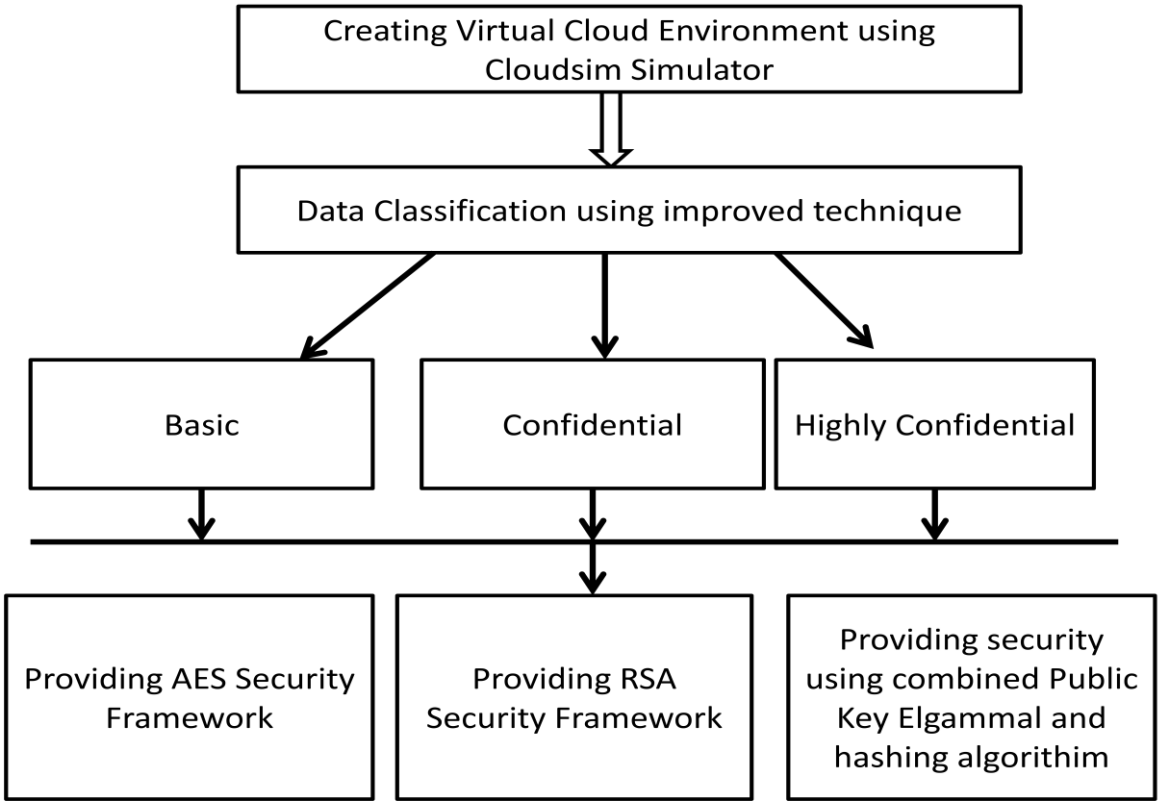

Fig 1: Proposed Methodology

4. RESULTS AND DISCUSSIONS

The proposed methodology is implemented with the help of Cloudsim and Net beans IDE 8.0. Cloudsim is the library that provides the simulation environment of cloud computing and also provide primary classes describing virtual machines, data centers, users and applications. The classification and encryption results have been illustrated in the following figures Figure 2, Figure 3, Figure 4, Figure 5 and Figure 6. And comparison between KNN with AES and Improved Naïve Bayes with different cryptographic techniques has been made in these figures

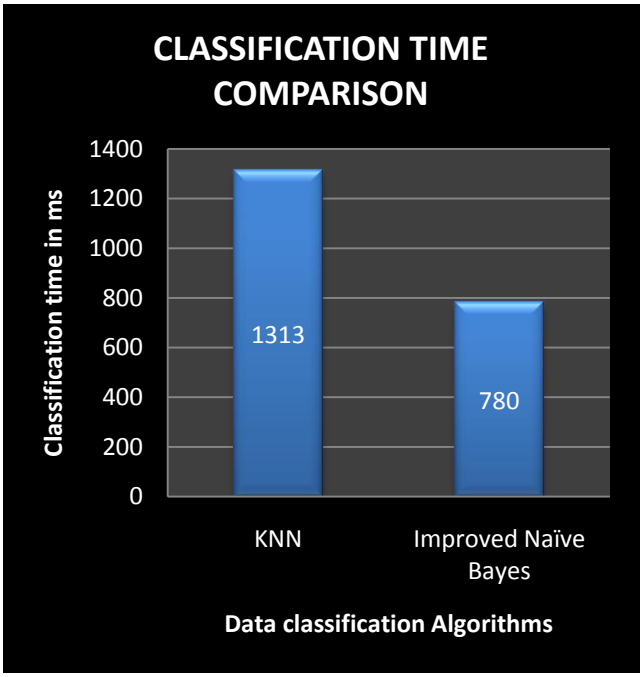

Fig.2.Performance analysis on the basis of classification time

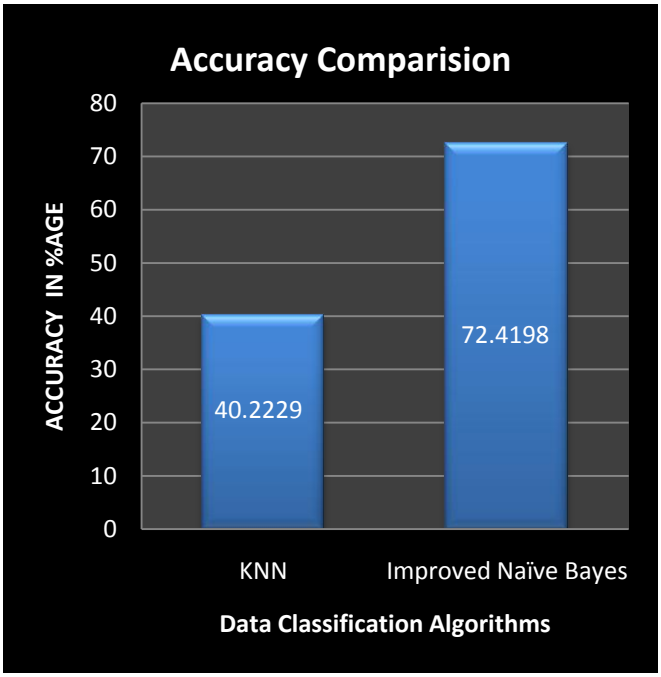

Fig.3.Performance analysis on the basis of accuracy

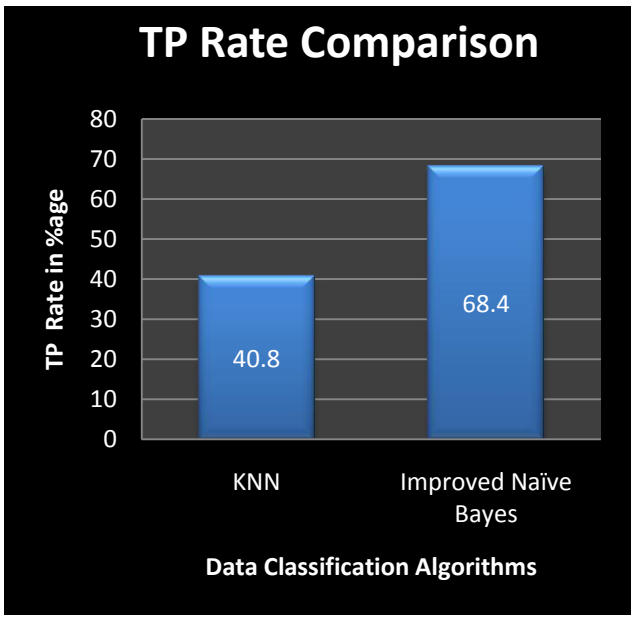

Fig.4.Performance analysis on the basis of TP Rate

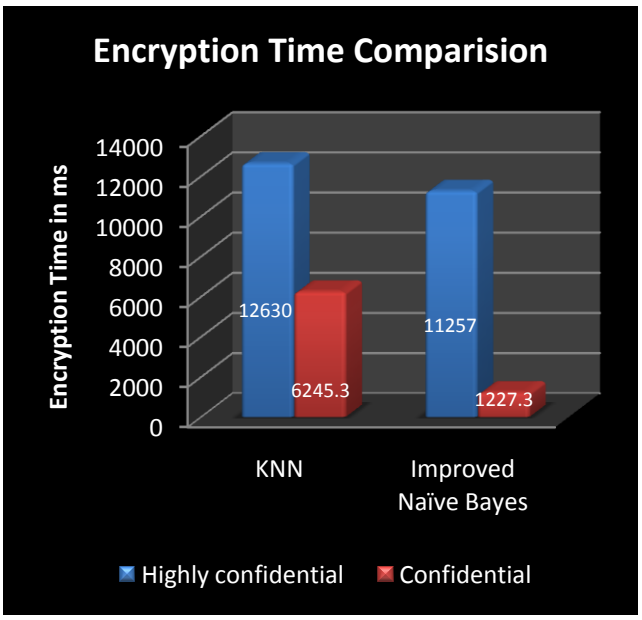

Fig.5.Performance analysis on the basis of Encryption Time

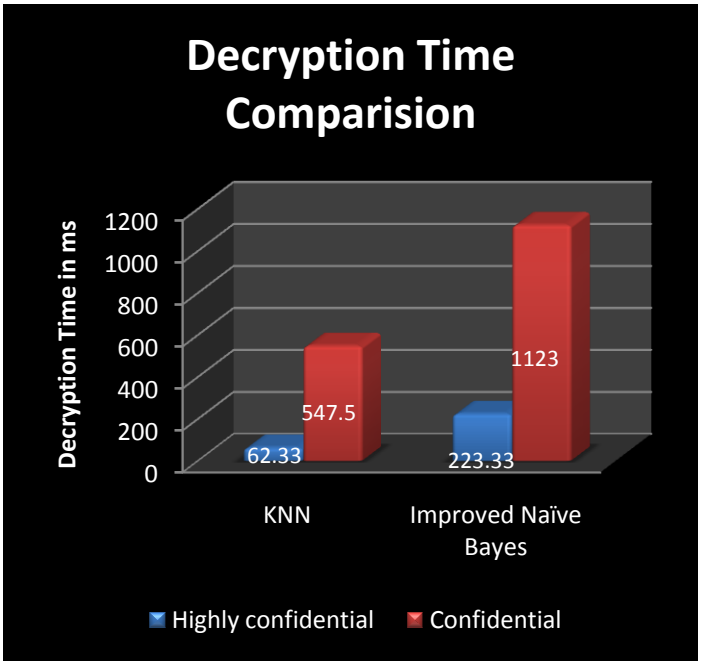

Fig 6. Performance Analysis Of Decryption time

The above figure shows the performance analysis of the proposed methodology with the previous method. It is clearly analyzed from the performance graphs that the proposed technique is better than the previous approach. Figure 2 shows the data classification time comparision and figure 3 shows accuracy comparison of data classification algorithms KNN and Improved Naïve Bayes Algorithm. KNN algorithm is having accuracy 40.2299% and improved naïve bayes is having 72.4198 % i.e. proposed algorithm has classified data more correctly. Similarly, figure 4.shows the true positive rate comparison between the KNN and Improved Naïve Bayes Similarly, figure 5 and 6 shows the more enhanced results in terms of encryption time and decryption time compared to the previous approach. Therefore, in order to reduce the encryption time on cloud data is classified according to its security needs using machine learning algorithms. From the above analysis it is shown that the proposed methodology performs betters in respect to Accuracy, data classification time, true positive rate, encryption time and decryption time.

5. CONCLUSIONS

In this research, a technique for data confidentiality in cloud environment is proposed. The focus of the research was to characterize the data taking into account the security prerequisites of the information that divides the data into basic, confidential and highly confidential data using improved machine learning algorithm. The fundamental contribution of this security model is data confidentiality and classification of data using machine learning classification approach. The classified confidential information is then encrypted using different cryptographic techniques and is stored in the cloud server .The proposed system has been simulated in a designed cloud simulation environment using cloud sim simulator. The results depict that the proposed technique is more relevant than storing the data without deciding the security needs of the data. Also, the results show that the improved naïve bayes technique works better than the K-NN classification technique in terms of both the accuracy ,classification time and TP rate and encryption time and decryption time also shows that the security is more enhanced in the proposed work.

In future, some more security requirements can be taken in account in order to take the classification decision using machine learning algorithm. Furthermore, to enhance the security at the authentication level, image sequencing passwords based on different themes will be used in order to avoid un-authorized access to the cloud environment. Authentication level security can be extended to multi-level authentication scheme so that each user will have different access permissions and roles. Availability of the encrypted data can also be improved in future.

6. ACKNOWLEDGMENTS

I might want to put on record my profound feeling of appreciation to Assistant Professor Mr. Rajeev kumar for his significant recommendations in my research work. I might want to thank every one of the general population whose consolation and support has made the satisfaction of this work possible.

7. REFERENCES

[1] Munwar ali zardari, Low Tang Jung, Nordin Zakaria," K-NN Classifier for Data Confidentiality in Cloud Computing", IEEE, pp.1-6, 2014.

[2] Almorsy, M., Grundy, J., & Ibrahim, A. S., "Collaboration- Based Cloud Computing Security Management Framework" IEEE conference of cloud computing, Washington (DC), pp. 364-371,2011.

[3] Song, D., E. Shi, I. Fischer and U. Shankar, "Cloud data protection for the masses", IEEE Computer. Soc., Vol. 45, Issue 1, pp.39-45, 2012

[4] Somani U, Lakhani K, Mundra M. Implementing digital signature with RSA encryption algorithm to enhance the Data Security of cloud in Cloud Computing. InParallel Distributed and Grid Computing (PDGC), 2010 1st International Conference on 2010 Oct 28 (pp. 211-216). IEEE.

[5] Dubey AK, Dubey AK, Namdev M, Shrivastava SS. Cloud-user security based on RSA and MD5 algorithm for resource attestation and sharing in java environment.

InSoftware Engineering (CONSEG), 2012 CSI Sixth International Conference on 2012 Sep 5 (pp. 1-8). IEEE.

[6] Yellamma P, Narasimham C, Sreenivas V. Data security in cloud using RSA. InComputing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on 2013 Jul 4 (pp. 1-6). IEEE.

[7] Rewagad P, Pawar Y. Use of digital signature with Diffie Hellman key exchange and AES encryption algorithm to enhance data security in cloud computing. InCommunication Systems and Network Technologies (CSNT), 2013 International Conference on 2013 Apr 6 (pp. 437-439). IEEE.

[8] Sinha N, Khreisat L. Cloud computing security, data, and performance issues. In2014 23rd Wireless and Optical Communication Conference (WOCC) 2014 May 9 (pp. 1-6). IEEE.

[9] Diwan V, Malhotra S, Jain R. Cloud security solutions: Comparison among various cryptographic algorithms. IJARCSSE, April. 2014 Apr.

[10] Zardari MA, Jung LT, Zakaria N. K-NN classifier for data confidentiality in cloud computing. InComputer and Information Sciences (ICCOINS), 2014 International Conference on 2014 Jun 3 (pp. 1-6). IEEE.

[11] Shaikh, Rizwana, and M. Sasikumar. "Data Classification for achieving Security in cloud computing." *Procedia Computer Science* 45 (Elsevier-2015): 493-498.

[12] Lo'aiTawalbeh NS, Raad S. Al-Qassas and Fahd AlDosari,"A Secure Cloud Computing Model based on Data Classification". InFirst International Workshop On Mobile Cloud Computing Systems, Management and Security (MCSMS-2015) 2015 (Vol. 52, pp. 1153-1158).