Journal of Engineering, Science and Sustainability

ISSN: 3107-474X (Online)

Vol. 1(2), Jul-Dec 2025, pp, 20-23

Journal homepage: *https://sprinpub.com/jess*

Research ***article***

Enhancing Machine Learning Model Security in Private Cloud Environments through Cryptographic Techniques

Ayesha Siddiqui[1], *Mohd Naved Ul Haq*[2*], *Mohd Nafees*[3]

[1-3]*Department of Computer Science and Engineering, Glocal University, India*

ARTICLE INFO

Keywords:

Machine learning, Cloud Computing, Cryptography Algorithms, Classifier

Article History:

Received: 11-08-2025

Revised: 26-11-2025

Accepted: 01-12-2025

Published: 11-12-2025

ABSTRACT

Cloud-based systems frequently encounter more serious security issues, especially when trying to secure stored information from harm. We present a model that acts as a barrier between users and cloud services, making use of modern techniques in machine learning for enhanced protection. Such a model is able to stop network attacks, identify different types of data and safeguard users' information by applying various encryption algorithms based on its relevance. The Random Forest and Decision Tree methods are very accurate at identifying attacks with a score of over 98% and Logistic Regression also reaches or surpasses 98%. Important classified data in the cloud is secured using methods such as Twofish, Blowfish and the Advanced Encryption Standard (AES).

Introduction

Machine learning, in particular, has transformed computing as more secure private clouds have emerged. Organizations are increasingly relying on private cloud solutions to protect sensitive data and maximize the benefits of machine learning. When we combine secure private clouds and machine learning, complex security issues arise, and cryptography becomes extremely important [1].

It is smart to apply cryptographic security to private cloud systems for machine learning to protect the increasing demand for confidentiality and safe storage of data. Big data and machine learning decisions can face serious security and privacy challenges. By using cryptographic techniques, we ensure that private ML data and models are not modified or accessed illegally.

This study explores the ways in which protocols for machine learning are connected and protected in private cloud systems. Through cryptographic security, this cooperation not only ensures machine learning continues to be a secret process but also gives stakeholders confidence in sharing important data in private clouds. We then look at the importance of protecting cryptographic security in secure private clouds and explain how vital it is to ensure machine learning projects can be done securely.

Private cloud helps businesses by allowing them to deploy their resources either over the internet to the public or privately within their internal systems. Many people prefer private cloud services offered by providers or organizations due to the greater benefits these offer. Its main features are self-service, the ability to scale, adaptability to different types of data and strong administration and control. The use of firewall services and internal hosting in the private cloud guarantees more security by letting only approved users have access to the data and preventing unauthorized use [2]. In recent research, a new way of classifying threats has been introduced [3]. The system uses machine learning algorithms, carried out using both supervised and unsupervised learning, to understand security challenges. You start by selecting the important model characteristics and group threats according to their differences, including those that affect all networks as well as those that are unique to the cloud environment.

Literature Review

Zecheng He et.al [4] outlines a DDoS attack detection system that relies on machine learning to protect cloud networks from attacks by anonymous virtual machines. By using statistics collected on the cloud server's hypervisor and virtual machines, the system manages to detect all DDoS traffic with an accuracy over 99.7%, making sure that the false positive rate stays low at 0.07% or even lower. Early on, the system blocks the outbound network packets, acting as the first intervention and keeping the cloud provider's reputation intact. Further developing the system with multiple machine learning methods, mainly aimed at strengthening unsupervised machine learning and increasing the tool's capabilities to detect several different DDoS attack attempts.

Ghassan et al [5] states that the use of cloud computing may have information privacy risks and offers a solution using RSA encryption and CHAP authentication. Using this method, both confidentiality and access control are ensured, making the method practical and successful. This approach deals with cloud security issues by using RSA encryption and an authentication protocol which prevents everyone but authorized users from accessing the information.

Mahmoud M. Sakr and his colleagues [6] present a solution to security problems in cloud computing by designing a NIDS that detects unusual activity. The system makes use of the SVM classifier and couples Binary-based Particle Swarm Optimization (BPSO) and Standard-based Particle Swarm Optimization (SPSO) for both choosing the correct features and setting optimal parameters. The NIDS is evaluated with the NSL-KDD dataset and is proven to highly accurately detect several types of attacks with very low false positives. In comparisons, Nipper is found to be more effective than other Intrusion Detection Systems (IDSs). It focuses on ways to set the parameters of the classification algorithm and use the best features for the network design.

* ***Corresponding Author***:

navedmuzakkir@gmail.com (M. N. U. Haq)

https://doi.org/10.55559/jess.v1i2.587

Gopal Krishna Shyam et.al [7] Summary: The paper looks at the problems related to security, focusing on data privacy, authentication and network security in cloud computing. While cloud technology helps businesses lower costs and be more convenient, it is not without risks such as data breaches and distributed attacks. The writer suggests using tools such as next-generation firewalls and control-based technologies to deal with such issues. The study covers both traditional and advanced artificial intelligence approaches and lists areas in artificial intelligence that still need to be explored. The paper points out the usefulness of such strategies and shows how machine learning is becoming more valuable in safeguarding the cloud.

In the research by Ibrahim S. I. Abuhaiba et.al [8], the objective was to enhance Arabic text document classification using four models with distinct algorithms. Initially, I applied rules that didn't vary, with majority voting between seven classifiers. This resulted in an accuracy of 95.3% after 836 seconds. The second model worked with stacking, resulting in very high accuracies of 99.2% for Naïve Bayes and 99.4% for linear regression, although it required more time to train. The third model adopted AdaBoost together with C4.5, resulting in a higher accuracy of 95.3% with 5 iterations and 99.5% with 10 iterations. The fourth model used the bagging algorithm with a Decision Tree, reaching an accuracy of 93.7% (based on 5 iterations).

99.4% (10 iterations). They performed better even though there were some issues with the models.

various classifiers are measured by their accuracy, precision, recall and F-measure scores. Moreover, incorporating alternative classifiers made stacked models faster to train and easier to use with larger data sets.

Xin Li et.al propose the LNNLS-KH algorithm which fixes the "dimensional disaster" problem in network intrusion detection by selecting features with high accuracy [9]. The algorithm uses linear nearest neighbour lasso optimization and creates a fitness function that counts the number of chosen features and the accuracy of the classification. Results from testing on NSL-KDD and CICIDS2017 suggest that the method is effective, as it reduces the number of features by 44% to 57.85% and performs more accurately than similar algorithms. LNNLS-KH's potential to avoid staying stuck in local optima is emphasized and the researchers suggest working on data balancing techniques and hybrid combinations for better performance in feature selection.

In their study, N. Pandeeswari et al. [10] introduced the Hypervisor Detector which identifies anomalies at the hypervisor layer in cloud systems. The system performs well in spotting both kinds of threats, using a combination of FCM-ANN and hybrid Fuzzy C-Means clustering

outsider attacks. The Hypervisor Detector, when tested on the DARPA KDD dataset with Naïve Bayes and Classic ANN, displays a high amount of accuracy and few false reports, most of all in detecting infrequent cyber-attacks. Including fuzzy clustering, training deep ANN sections with the results and a Fuzzy aggregation module in the FCM-ANN model, boosts the learning ability of the ANN. This demonstrates that it is highly effective at spotting unauthorized access to cloud networks.

Basel Saleh Al-Attab et al. [11] Given the rising security problems with cloud computing, the paper presents a new way to secure data by using hybrid encryption to tackle key vulnerabilities, fast processing and speed. The algorithm is designed to stress data security by putting emphasis on cryptography and it uses asymmetric key, secret sharing and key exchange. By using both approaches, the goal is to make data more secure in the cloud. Increasing the use of security algorithms in managing and handling data systems. Developing the algorithm to specifically satisfy cloud computing requirements, putting priority on quick and efficient handling of data

This research [12] strongly suggests that having strong access controls and including MFA is necessary for protecting data saved in the cloud. The study points out that traditional MFA systems are overly inconvenient and that cloud servers can be insecure which is why a new trust model is created to change the authentication procedure according to the device in use. Using biometrics on public devices and one-factor authentication for private devices, the trust-based MFA gets better results. With the introduction of the MACA system, both multifactor authenticity and privacy are protected, since it combines a password and a hybrid user profile, making the whole process more efficient and resourceful as indicated by the evaluation results.

In [13], Deukjo Hong et.al put forth the LEA block cipher which is suitable for software encryption and comes with 128-bit block sizes as well as key sizes of 128, 192 or 256 bits. Results show that experiments outperform AES on several platforms and LEA is able to maintain a reduced code size. Its software works well in practice with effective throughput and strong defense against various cipher attacks. While the research has performed very well, the authors suggest that there are further improvements that could be made.

This research by Dursun Delen et.al [14] discusses how text mining is important for understanding large volumes of unstructured data. Using text mining on abstracts from important Management Information Systems journals lets us analyze trends and group together similar research topics. Automated ways of managing data are becoming more valuable because of the large amount of information being processed. It is expected that future efforts will develop algorithms that can handle synonyms and the context of words. It is suggested to use both data mining and text mining for a complete way to extract knowledge from any type of information.

The study conducted by Hasan Kamel et.al [15] looks at SDN security risks and proposes using machine learning to identify and classify DDoS attacks in such networks. When I used Dataset with UDP, TCP and ICMP protocols, my proposed Evolutionary Decision Tree (EDT) model, optimized with the Genetic Algorithm (GA), achieved an impressive classification accuracy of 99.46%. Thanks to selecting the right hyperparameters, the GA-enhanced model achieved strong performance in separating normal and attack traffic.

Singh et.al [16] This work proposes a way to make sure that data is securely exchanged between federated cloud entities by performing mutual authentication. Using Elliptic Curve Cryptography and Schnorr's signature scheme, the method ensures secure exchange of messages. It also includes a real-time threat detection system using an ensemble Voting Classifier based on machine learning. With the help of Canadian Institute for Cybersecurity Datasets and ProVerif tool, the protocol is proven to be efficient in terms of security and the cost of communication. Thanks to its resilience against attacks, the lightweight protocol provides anonymity to users and protects session keys, so it is safe to use for online data exchange in a multi-cloud setting.

In their study, Abbas Jasem Altamemi et.al [17] use machine learning to quickly spot DDoS attacks in Software Defined Networking (SDN). The algorithms examined in the study are Decision Tree (DT), Naïve Bayes (NB) and Logistic Regression (LR) which are used for classifying DDoS attacks in SDN. The

the proposed system outperforms others in accuracy, reaching 99.90% for DT. By using machine learning, SDN is able to respond swiftly to DDoS attacks. Finding ways to make the system more optimized, faster in responding and able to grow larger without costing too much.

Hassan and team [18] This paper presents an approach for detecting and mitigating DDoS attacks in IoT networks using Fog Computing. Being close to IoT devices, Fog Computing can detect attacks quickly and accurately. It uses the method of traffic randomness measurements along with the KNN approach in machine learning. The accuracy of the system is very high when it comes to detecting faces. The study reported success in TCP attacks at 100%, 98.79% for UDP attacks and 100% for ICMP attacks. It enables fast action against an assault, lessening the burden on the hardware. Introduce different types of DDoS attacks and update physical detection capabilities for Fog Computing.

Proposed Work

The proposed model includes adding a third-party layer between users and the private cloud, as illustrated in Figure 1. Using ML, the model determines if a user's request is normal or appears to indicate a malicious attack. Later, normal requests continue moving forward and harmful ones are quickly detected and stopped. The system is configured using the BBC News dataset, a popular and effective resource for this type of work. Here, requests generated by users are studied and their data on the cloud is arranged according to how important it is—high confidentiality, confidentiality or just basic. After final classification, data is encrypted with suitable algorithms.

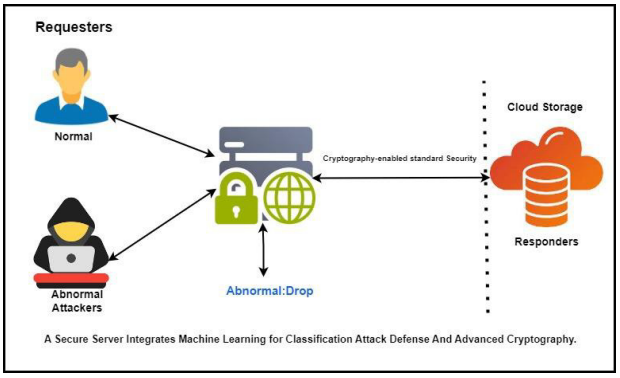

Figure-1: *Third-Party Model Situated Between Users and The Private Cloud*

The security system includes powerful algorithms such as AES-256, Blowfish and Twofish. To handle datasets with varying privacy levels, the team decided to use BBC News as their source. This data is used to train the classifier which deals with different topics such as Health, Science, Environment, Fashion and technology. The approach suggested for unclassified requests in Figure 2 relies on using request classification data along with machine learning tools.

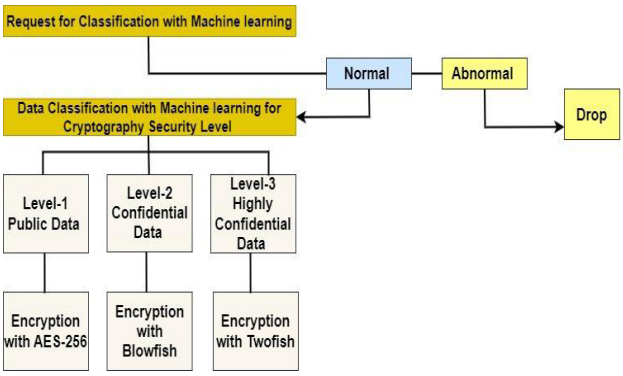

Figure-2: *Data Classification Through Machine Learning Techniques*

- **Using machine learning to organize data into different classes or categories**

The system management of text data from the BBC News dataset consists of four major sub-steps: i) Obtaining the data from BBC, ii) preparing the data, iii) selecting important features and iv) training the model to perform classification. BBC news data we are using comes from the British Broadcasting Corporation (BBC) and consists of raw text files available here: https://www.kaggle.com/search?q=bbc+news. The files downloaded from the BBC news website document events happening around the country during the chosen time span. 2005 to 2006. The dataset is divided into five folders, called Health, Science, Environment, Fashion and Technology, each filled with news articles connected to that class label.

- **Pre-processing**

Using tokenization and drop of stop words improves the quality of input data substantially. It is important for this step because it shapes and prepares the text for the next process. According to the BBC dataset, the documents for each article are stored in separate folders, with the category shown in the folder name (e.g., 'technology' for topics on business). Then, every category is identified by a specific number (0 for Health, 1 for Science, 2 for Environment, 3 for Fashion, and 4 for Technology). In Table 1, we can see how the samples are divided among the different categories.

Category	ID Number	Number of Items	Cryptography Algorithm
Health	0	450	Twofish
Science	1	320	Blowfish
Environment	2	390	Twofish
Fashion	3	480	Nothing
Technology	4	350	AES

Table 1. *Proposed ategories dataset*

The classification phase is detailed in Figure 4 with data pre-processing, feature extraction, and the classifier's training-testing steps. The primary reasons for pre-processing datasets are: 1) to make the dataset more manageable for data analysis, and 2) to increase the method's ability to analyze the dataset. It mainly helps by shrinking the size of a dataset by ignoring unnecessary features for categorization.

$$\text{Weight}\ (x,y) = \frac{tf(x,y)}{\max tf(y)} \times \log\left(\frac{c}{zx} + 1\right)$$

Its address is:

The term x appears in the document y times and is measured by tf(x,y).

The most frequent term in document y is displayed by max_tf(y). c is the number denoting all documents in the collection.

iv. zx means the number of documents having the term x.

The weight value is found by making tf(x,y) proportional to max_tf(y) for the given document. The more often the term x appears in the documents, the greater the power of the logarithmic function.

Algorithm Steps

Input

i. The term frequency of term x in document y is denoted as tf(x, y).

ii. The maximum term frequency in document y is expressed by max_tf(y). iii. The count of documents in the text collection is identified as c.

iv. The count of documents where term x occurs is labeled as zx.

Calculate Weight:

i. Calculate the normalized term frequency: tf (x , y).
max_tf(y)

ii. Apply the logarithmic transformation: log(c/zx+1).

iii. Multiply the two results to obtain the term weight.

Output:

The computed term weight for term x in document y.

As previously stated, classification features are implemented in two phases:

a. In this stage, each request is examined and checked to tell if it is standard or not.

b. During this phase, we organize and sort all concealed stories by their category.

Skills for data pre-processing are:

a. A lexical analyzer first divides the text into smaller units such as words, letters or entire sentences.

b. A way to process data that uses NLP techniques and stops filtering unnecessary words is known as Stop Word Filtering.

technology. The purpose is to get rid of unnecessary and meaningless terms to slim down the data set using stop words. This simplifies how remaining keywords are found using automated approaches.

d. Scaling: In this part, each word is put in the correct scale (e.g., letters to lowercase) and list is sorted. This predictability brings a uniform format and makes it possible to avoid using capital letters or numbers.

d. Stemming is applied to improve the search process, and Snowball stemming is the method used. It removes the endings or beginning letters from words to get their basic form.

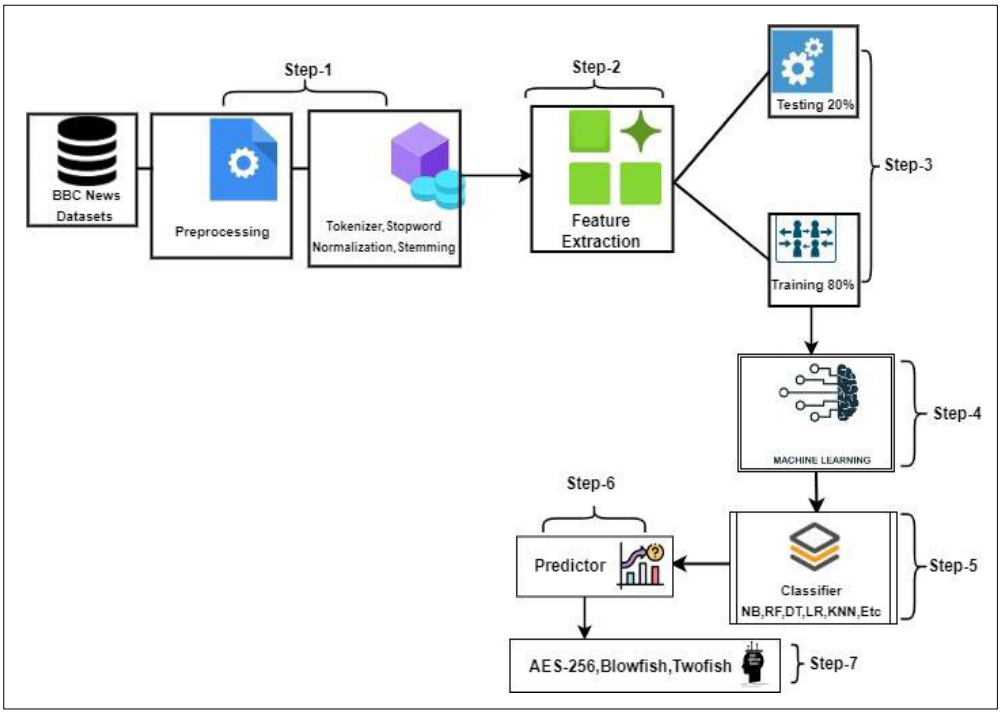

Figure-4 *Proposed Model*

Methodology

Machine learning (ML) methods are used in this methodology, which consists of two main phases. First, it consists of identifying the requests meant for the private cloud. It identifies which transactions are normal and which may be suspicious. Training for the classifier uses the "BBC News" dataset. In step two, the data in the personal cloud is organized with higher importance, such as Top Secret, being labeled as most sensitive; Confidential is next in importance; and the remaining data is labeled as Basic Data. AES-256, Blowfish, and Twofish are used to encrypt the data categories before they are sent to a server. Read on to learn about the traits, functions, and purposes of machine learning algorithms.

The fact that Naive Bayes is a probabilistic algorithm makes it effective for tasks like text classification and spam filtering. The performance of NB is impressive, since it can make predictions with only moderate understanding of the input. The reason it is popular is that it can deal well with missing values and is efficient. NB is safer from overfitting because it makes simple assumptions and uses probabilities.

RF is known for being effective and reliable because it is an ensemble learning algorithm. RF works well with data because it uses multiple decision trees to deal with non-linear patterns. The ensemble nature of the model tends to reduce the risk of overfitting, though it does require proper tuning. RF is good for solving many classification and regression tasks and is relatively quick to scale.

A feature of DT is its tree structure, known for being easy to interpret and suited to making decisions. In handling non-linearity, DT takes input features and decides whether they belong to one class or the other. But, since overfitting is a risk, the depth and yardstick for splitting need to be set carefully. In cases where interpretability matters a lot, such as binary classification and issues with decision trees, DT is a powerful tool [24].

Logistic Regression, despite the name, is primarily used to perform binary classification and is a linear model. Estimates of probabilities by LR make use of the logistic function for its interpretable results. Thanks to LR's versatility, it is not limited to using it just for binary classification, and can help comprehend how each feature impacts the outcome [25].

Tabel-2 *Compression Tables of Ml Algorithms*

Aspect	NB [2]	RF [3]	DT [4]	LR [5]	KNN [6]
Type	Probabilistic	Ensemble	Decision Tree	Linear Model	Instance-based
Interpretability	Moderate	Moderate to Low	High	Low	Low
Training Speed	Fast (no iterative training)	Moderate to Fast	Fast	Fast	No training phase
Predictive Speed	Fast	Fast	Fast	Fast	Slow (distance alculation for all points)
Scalability	High	Moderate to Fast	Moderate to High	High	Low to Moderate
Feature Scaling	Not sensitive	Not required	Not required	Required	Sensitive (distance-based)
Handling Missing Values	Handles missing values naturally	Handles missing values naturally	Handles missing values naturally	Requires imputation or elimination	Requires imputation or elimination
Non-linearity	Not applicable (Assumes Independence)	Handles non-linearity well	Handles non-linearity well	Limited	Can capture non-linear Patterns
Robust to Outliers	Not sensitive	Robust	Sensitive	Sensitive	Sensitive
Hyperparameter Tuning	Laplace smoothing (for NB)	Number of trees, depth, features	Depth, splitting criteria	Regularization, possibly feature scaling	Number of neighbours, distance metric
Overfitting	Less prone	Less prone (due to ensemble)	Prone without proper tuning	Less prone	Prone without proper tuning
Use Cases	Text classification, spam filtering	General- purpose classification, regression	Classification, regression	Binary classification, linear problems	Classification, regression, clustering

KNN

KNN classifies data points using the majority class of similar points close to them. Even though KNN gives simple and useful results, its predictive speed is slower, which is noticeable in fast processing situations. KNN is commonly applied in classification, regression, and clustering tasks, as it balances accuracy with understanding [23].

Advanced Encryption Standard

The AES, or Advanced Encryption Standard, is a symmetric key encryption that is commonly applied in different fields. The key sizes offered are 128, 192, and 256 bits, ensuring the highest level of privacy. AES offers high protection with its 10, 12, or 14 round processes. This innovation employs the substitution method.

There is an SPN structure, key expansion, the addition of S-boxes, and key whitening. When introduced in 2001, AES became the preferred choice for encrypting data quickly and completely.

Twofish

Twofish, another symmetric key algorithm, is created for everyday use and experts recommend it.

particularly for constrained environments. The mode works with key lengths of 128, 192, and 256 bits and employs Feistel structure, key expansion, S-boxes, and key whitening. Twofish was developed in 1998 and excels at offering a secure and fast option for different uses.

Blowfish

Blowfish was released in 1993 and ensures strong security for most applications, with special advantages when used in software. To operate, Blowfish adopts a 64-bit fixed key size, Feistel structure, key expansion, and S-boxes, while also including key whitening. It provides medium to high-security strength when locked with 16 rounds. Even though it is not the quickest, Blowfish is appropriate when balancing security and the need for efficiency [26].

Table-3 *Cryptographic Algorithms Compression Tables*

Factor	AES	Twofish	Blowfish
Algorithm Type	Symmetric	Symmetric	Symmetric
Key Size (bits)	128,192,256	128,192,256	64
Rounds	10,12,14	16	16
Feistel Structure	No	No	Yes
Key Expansion	Yes	Yes	Yes
S-Boxes or Substitution Permutation Networks (SPNs)	Yes	Yes	Yes
Key Whitening	Yes	Yes	Yes
Security Strength	High	Very High	Medium to High
Speed	Fast	Moderate to Fast	Moderate
Confidentiality	Excellent	Excellent	Good
Use Cases	General Purpose, widely adopted	General Purpose, recommended for constrained environments	General Purpose, particularly for software implementations
Year Introduced	2001	1998	1993

Result

In the proposed strategy, the focus is on configuring the server as an intermediary, the major objective of which is to strengthen the security measures between these organizations. Through the utilization of machine learning and encryption techniques, a system has been designed to enhance cloud security and identify potential threats. The "BBC News" dataset is employed for the purpose of classification, training the classifier with privately stored data within a cloud. The evaluation phase includes testing various algorithms such as Naive Bayes, Logistic Regression, K-Nearest Neighbours, Random Forest, and Decision Trees. In the suggested machine learning algorithm-based classification process, two primary phases exist: data training and testing. Based on specific criteria, comparative analysis is conducted using performance metrics like accuracy, precision, recall, and F1- score [**9**].

Tabel-4 P*erformance Parameters with Equation*

Performance Parameter	Equation
Accuracy	True Positives + True Negatives True Positives + Fake Positives + Fake Negatives + True Negatives
Precision	True Positives True Positives + Fake Positives
Recall	True Positives True Positives + Fake Positives
F1-Score	2 × Precision × Recall Precision + Recall
DR	True Positives True Positives + Fake Negatives
FAR	Fake Positives Fake Positives + True Negatives
ERR	b + c /a+b+c+d

i. TP means that the model correctly identified something positive.
ii. ii. A false positive (FP) is a case where the model thinks something is positive when it should be considered negative.
iii. When false negatives occur, the model says something is negative even though it should be positive.
iv. TN stands for false negatives, which are correctly predicted as negative by the model [19, 20].

To find accuracy, the model is measured by counting its number of correct positives and negatives. It provides an idea of how accurately the system differentiates instances [21].

The main idea behind precision is that it shows how accurate the predictions of positive results are among all positive results identified. It is used to measure how well the model is at picking out positives. [10].

Recall is used to assess how correctly a model can find all cases that belong to a specific class. It assesses the number of actual positives the model correctly detects [11].

F1-Score is a harmonic mean of precision and recall; it helps identify how well a result considers both errors at the same time. In situations where one class is much larger than the other, this proves especially helpful [12].

The False Alarm Rate (FAR) shows how often a detector gives a wrong alert versus the number of real negatives. This brings forward the tendency of the model to incorrectly label negative instances as positive [16].

Error Rate (ERR) presents a comprehensive evaluation of classification errors, considering both false positives and false negatives relative to the total number of instances. It serves as a complement to accuracy [18].

Table 2 and Figure 5 provide an outcome of the algorithms that were used. When compared to other algorithms, the DT and RF algorithms perform better in terms of accuracy and detection rate, building algorithm, as well as other areas; but, since the LR method relies on randomization, it is thought to be the fastest. This study suggests LR's optimal efficiency for examined datasets may be attributed to its quick training model and straightforward implementation.

Table -5 *Result Table of Algorithms*

Aspect	Naive Bayes	Random Forest	Decision Tree	Logistic Regression	KNNK- Nearest Neighbours
Precision	0.79	0.96	0.94	0.88	0.91
Accuracy	0.75	0.93	0.92	0.98	0.88
Recall	0.78	0.95	0.93	0.87	0.90
F1-Score	0.78	0.96	0.94	0.89	0.92
FAR	0.41	0.03	0.05	0.11	0.13
DR (Detection Rate)	0.92	0.97	0.98	0.86	0.93

In short, Logistic Regression simplifies running time by converting the features of the dataset into groups of related attributes. It then ranks the clusters based on the similarities among them. LR needs time to consider all the data at once, but it quickly finds accurate results. The visual representation of the results from the proposed phase is shown in Figure 5.

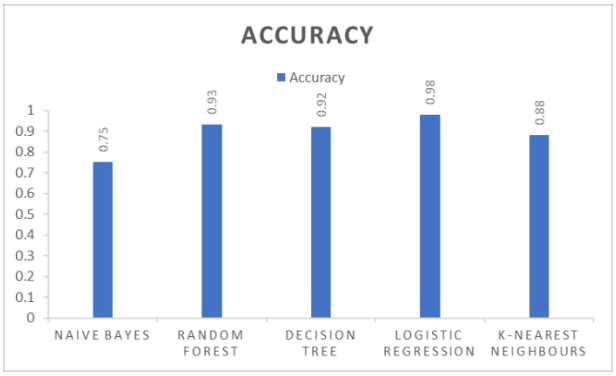

Figure 5 *Accuracy Graph for all Algorithms*

Conclusion

The paper presents a secure cloud architecture that makes use of machine learning and cryptography. It uses several levels of encryption, protecting sensitive data first while keeping the use of resources at a minimum for regular data. Using the BBC classification dataset, it was observed that the LR algorithm among several other ML models was the most effective in identifying attack traffic from normal network traffic. In the second phase, the LR model accurately classifies different data types with a high precision of 98%.

References

[1] Jabbar, A. A., & Bhaya, W. S. (2023). Security of private cloud using machine learning and cryptography. Bulletin of Electrical Engineering and Informatics, 12(1), 561-569.

[2] Goyal, S. (2014). Public vs private vs hybrid vs community-cloud computing: a critical review. International Journal of Computer Networks & Information Security, 6, 20.

[3] Masetic, Z., Hajdarevic, K., & Dogru, N. (2017). Cloud computing threats classification model based on the detection feasibility of machine learning algorithms. 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1314–1318. doi: 10.23919/MIPRO.2017.7973626.

[4] He, Z., Zhang, T., & Lee, R. B. (2017). Machine Learning Based DDoS Attack Detection from Source Side in Cloud. Proceedings of the 4th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud 2017), 114–120. doi: 10.1109/CSCloud.2017.58.

[5] Mahmood, A. L. G. S. (2017). Data Security Protection in Cloud Computing by using Encryption. Kirkuk University Journal/Scientific Studies, 12(4), 849–1992.

[6] Sakr, M. M., Tawfeeq, M. A., & El-Sisi, A. B. (2019). Network Intrusion Detection System based PSO- SVM for Cloud Computing. International Journal of Computer Networks & Information Security, 11(3), 22–29. doi: 10.5815/ijcnis.2019.03.04.

[7] Shyam, G. K., & Doddi, S. (2019). Achieving Cloud Security Solutions through Machine and Non- Machine Learning Techniques: A Survey. Journal of Engineering Science and Technology Review, 12(3).

[8] Abuhaiba, I. S. I., & Dawoud, H. M. (2017). Combining different approaches to improve Arabic text documents classification. International Journal of Intelligent Systems and Applications, 9(4), 39. doi: 10.5815/ijisa.2017.04.05.

[9] Li, X., Yi, P., Wei, W., Jiang, Y., & Tian, L. (2021). LNNLS-KH: A Feature Selection Method for Network Intrusion Detection. Security and Communication Networks, 2021, 8830431. doi: 10.1155/2021/8830431.

[10] Pandeeswari, N., & Kumar, G. (2016). Anomaly detection system in cloud environment using fuzzy clustering-based ANN. Mobile Networks and Applications, 21(3), 494–505. doi: 10.1007/s11036-015-0644-x.

[11] Al-Attab, B. S., & Fadewar, H. S. (2018). Hybrid data encryption technique for data security in cloud computing. Sinhgad Institute of Management and Computer Applications, 221-224.

[12] Anakath, A. S., Rajakumar, S., & Ambika, S. (2019). Privacy preserving multi-factor authentication using trust management. Cluster Computing, 22(5), 10817–10823. doi: 10.1007/s10586-017-1181-0.

[13] Hong, D., Lee, J.-K., Kim, D.-C., Kwon, D., Ryu, K. H., & Lee, D.-G. (2013). LEA: A 128-bit block cipher for fast encryption on common processors. International Workshop on Information Security Applications, 3–27. doi: 10.1007/978-3-319-05149-9_1.

[14] Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. Expert Systems with Applications, 34(3), 1707–1720. doi: 10.1016/j.eswa.2007.01.035.

[15] Kamel, H., & Abdullah, M. Z. (2022). Distributed denial of service attacks detection for software- defined networks based on evolutionary decision tree model. Bulletin of Electrical Engineering and Informatics, 11(4), 2322–2330. doi: 10.11591/eei.v11i4.3835.

[16] Singh, A. K., & Saxena, D. (2022). A cryptography and machine learning-based authentication for secure data-sharing in a federated cloud services environment. Journal of Applied Security Research, 17(3), 385–412.

[17] Altamemi, A. J., Abdulhassan, A., & Obeis, N. T. (2022). DDoS attack detection in software-defined networking controller using

machine learning techniques. Bulletin of Electrical Engineering and Informatics, 11(5), 2836–2844. doi: 10.11591/eei.v11i5.4155.

[18] Hassan, K. F., & Manaa, M. E. (2022). Detection and mitigation of DDoS attacks in the Internet of Things using a fog computing hybrid approach. Bulletin of Electrical Engineering and Informatics, 11(3), 1604–1613. doi: 10.11591/eei.v11i3.3643.

[19] Kumar, V., Sinha, D., Das, A. K., Pandey, S. C., & Goswami, R. T. (2020). An integrated rule-based intrusion detection system: analysis on UNSW-NB15 data set and the real-time online dataset. Cluster Computing, 23(2), 1397–1418. doi: 10.1007/s10586-019-03008-x.

[20] Vora, U., Mahato, J., Dasgupta, H., Kumar, A., & Ghosh, S. K. (2021). Machine Learning–Based Security in Cloud Database—A Survey. Machine Learning Techniques and Analysis for Cloud Security, 239–269. doi: 10.1002/9781119764113.ch12.

[21] Uçtu, G., Alkan, M., Doğru, İ. A., & Dörterler, M. (2021). A suggested testbed to evaluate multicast network and threat prevention performance of Next Generation Firewalls. Future Generation Computer Systems, 124, 56–67. doi: 10.1016/j.future.2021.05.013.

[22] Bhavani, D. D., Vasavi, A., & Keshava P. T. (2016). Machine Learning: A Critical Review of Classification Techniques. International Journal of Advanced Research in Computer and Communication Engineering, 5(3), 22-28.

[23] Narudin, F. A., et al. (2016). Evaluation of machine learning classifiers for mobile malware detection. Soft Computing, 20, 343-357.

[24] Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. Artificial Intelligence Review, 56(5), 4765–4800.

[25] Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. Procedia Computer Science, 89, 117–123.

[26] Ul Haq, M. N., & Kumar, N. (2021). A novel data classification-based scheme for cloud data security using various cryptographic algorithms. International Review of Applied Sciences and Engineering.