Article

Enhancing Industrial Cyber Security, Focusing on Formulating a Practical Strategy for Making Predictions Through Machine Learning Tools in Cloud Computing Environment

Zaheer Abbas[1] and Seunghwan Myeong [2,*]

[1] Center of Security Convergence & eGovernance, Inha University, Incheon 22212, Korea; snoj215@gmail.com
[2] Department of Public Administration, Inha University, Incheon 22212, Korea
* Correspondence: shmyeong@inha.ac.kr

Abstract: Cloud computing has revolutionized how industries store, process, and access data. However, the increasing adoption of cloud technology has also raised concerns regarding data security. Machine learning (ML) is a promising technique to enhance cloud computing security. This paper focuses on utilizing ML techniques (Support Vector Machine, XGBoost, and Artificial Neural Networks) to progress cloud computing security in the industry. The selection of 11 important features for the ML study satisfies the study's objectives. This study focused on identifying gaps in utilizing ML techniques in cloud cyber security. Moreover. this study aims at developing a practical strategy for predicting the employment of machine learning in an Industrial Cloud environment regarding trust and privacy issues. The efficiency of the employed models is assessed by applying validation matrices of Precision, Accuracy, Recall values, F1 score, R.O.C. curves, and Confusion matrix. The results demonstrated that the X.G.B. model outperformed in terms of all the matrices with an Accuracy of 97.50 %, 97.60 % Precision, 97.60 % Recall values, and 97.50 % F1 score. This research highlights the potential of ML algorithms in enhancing cloud computing security for industries. It emphasizes the need for continued research and development to create more advanced and efficient security solutions for cloud computing.

Keywords: Cloud security; Cloud computing; machine learning; industrial cyber security

1. Introduction

Modern organizations have embraced cloud computing as a necessary component because it allows them to store, access, and analyze data effectively [1]. However, data security concerns also arise with the continuous use of cloud technology. Cloud computing security is the methods and tools used to safeguard information kept on cloud servers against unauthorized access, theft, and other security risks. Industries are a major target for cyberattacks because they store tremendous amounts of confidential data on cloud servers, including financial information, client data, and scientific property [2]. Therefore, it is essential to enhance cloud computing security to save valuable data from unauthorized access and theft. ML has been recognized as a promising method for enhancing cloud computing security. Different algorithms can examine huge data capacities, identify arrangements, and learn from the data to identify potential security threats. By using machine learning, industries can enhance their security measures and reduce the risk of data breaches [3, 4].One of the significant challenges of securing cloud environments is the lack of control over the physical infrastructure [5, 6]. The physical infrastructure, which consists of servers, storage units, and network devices, must be kept up by the cloud package providers. Therefore, to ensure that the data kept on cloud servers are secure, trust between the consumer and the cloud service provider is crucial [7]. Another challenge of securing cloud environments is the complexity of the system. Cloud

environments involve multiple layers of hardware, software, and networking components, making identifying and addressing security threats difficult [8]. Cloud computing has become increasingly popular for its scalability, flexibility, and cost-saving benefits. However, it also presents several vulnerabilities that can threaten the security and integrity of data and applications[9]. Some major vulnerabilities in cloud computing include data breaches, insider threats, malware attacks, denial of service (DoS) attacks, insecure APIs, shared infrastructure risks, lack of visibility and control, compliance and legal risks, data loss and data leakage, and lack of encryption[10]. Sensitive data kept in the cloud can be stolen by unauthorized parties frequently due to poor access controls, ineffective authentication procedures, or defects in the infrastructure or software of cloud service providers (C.S.P.s) [11]. Insider threats are risks posed by staff members or authorized users with access to protected cloud assets. These risks can include deliberate or accidental activities that jeopardize the confidentiality of the cloud environment [12]. Malware attacks, including viruses, worms, and ransomware, can infect cloud systems and compromise data integrity and confidentiality [13]. DoS attacks aim to degrade cloud services by overwhelming the cloud infrastructure or applications with excessive traffic or requests. Attackers may utilize insecure APIs, which are required to connect cloud services to other programs or systems, to obtain unauthorized access to or change data stored in the cloud [14]. Shared infrastructure risks arise from cloud services being built on shared infrastructure, where multiple users and applications share the same physical resources. Vulnerabilities in the underlying infrastructure, such as hypervisors or virtualization layers, can lead to cross-tenant data breaches or unauthorized access [15]. Lack of visibility and control in cloud environments can make it challenging for organizations to have full control and visibility into their security posture. Compliance and legal risks may arise from data storage in different geographic locations, subject to different legal jurisdictions and data protection regulations, making compliance with industry-specific regulations challenging [16]. Data loss and data leakage can occur due to accidental deletion, hardware failures, misconfigurations in the cloud environment, unauthorized data access, or sharing of data with unintended recipients. Lack of encryption can leave data transmitted or stored in the cloud vulnerable to interception or unauthorized access [17, 18]. Traditional security measures, such as firewalls and antivirus software, cannot protect cloud environments from sophisticated cyber-attacks [12, 19]. By using machine learning, industries can enhance their security measures and reduce the risk of data breaches [20]. The potential of machine learning to boost threats identification and responses is one of the main benefits of employing it for cloud security [21]. Traditional security measures such as firewalls and antivirus software are reactive and only respond to known threats. Machine learning, on the other hand, can identify patterns in data that may indicate a threat, even if the threat is not yet known. Based on past data, ML algorithms are generated to find designs that point to security vulnerabilities [22, 23]. A machine learning algorithm can be generated on network traffic data to identify behavioral patterns indicative of a cyber-attack. Once trained, the algorithm may track traffic on the network in real time and notify security staff of any unusual activity patterns. An additional advantage of utilizing machine models for cloud security is its ability to automate certain security tasks [23, 24]. Different machine learning algorithms can be recycled to automatically classify and prioritize security alerts, reducing the workload on security personnel. It can free security personnel to concentrate on more complex security jobs requiring human expertise. Machine learning can enhance access control and identity management [25]. By examining user behavior patterns, machine learning algorithms can identify out-of-the-ordinary activity that can point to unauthorized access. It can help businesses prevent unauthorized access attempts before they can cause any damage. One of the challenges of employing these models for cloud security is the lack of transparency in how machine learning algorithms make decisions [26]. Because machine learning algorithms may be complicated and tricky to read, security employees may find it difficult to comprehend why a certain warning was created. [27]. Businesses can use techniques such as explainable A.I. to address this challenge by making machine learning algorithms more transparent and interpretable. Another challenge is balancing security with usability [28, 29]. While security is important, businesses must also ensure that their cloud services are easy to use and do not create unnecessary friction for users. Machine learning can help businesses strike this balance by

automating certain security tasks and making security alerts more targeted and actionable [24]. Machine learning has the potential to significantly improve cloud computing security in the industry [30]. By improving threat detection and response, automating security tasks, and enhancing access control and identity management, machine learning can help businesses reduce the risk of data breaches and other security incidents. However, businesses must be aware of the challenges associated with utilizing ML techniques for cloud safety and invest the necessary resources to overcome these challenges [31].

2. Literature Review

Various approaches have been made in the past in which machine learning technology was utilized to enhance cloud computing security. A study examined the ML models to identify and prevent insider threats in the cloud. The study used a dataset of access logs from a cloud platform to sequence a machine learning model to categorize irregular performance indicative of an insider threat. According to the study, the ML model may produce results with a high degree of accuracy in detecting insider threats, demonstrating the potential of machine learning for improving cloud security [41]. Another study explored the use of ML algorithms for improving cloud network security. The study used a dataset of network traffic logs from a cloud platform to develop a machine-learning model to recognize patterns of conduct indicative of a cyber-attack [42]. Marwan et al. used SVM and F.C.M. to improve safety in healthcare departments. Their main concern was preventing unauthorized access to medical records, and the suggested approach provides the opportunity to process RGB colorful photos safely using cloud services [34]. Linear regression and SVM boost cyber security by performing the static validation of cloud user action to spot the predefined danger. These security measures are limited in their function because they are static. Subramanian et al. [35] designed the novel security ML model CNN, offering quick and automated solutions to improve security. It offered suggestions for integrating detailed algorithms for securing data throughout all cloud apps instead of solely recognizing sensitive data trends. Mohammad et al. proposed the ML-assisted cloud computing module to enhance safety, and this model exhibited a performance of 95.2 %. It also increased the data transmission rates to 96.4 % [43]. The security of the hybrid cloud was also tried to increase by using the ML approach, which enhanced $C_{4.5}$ combined with a deduplication algorithm and access control mechanism [36]. Tabassum et al. utilized the Neuro-fuzzy ML algorithm to address the cloud cybersecurity challenges. The ANFIS-dependent variables were largely made to identify the anomalies related to cloud privacy [44]. Linear regression (L.R.) and Random Forest (R.F.) were also utilized in the past to detect the threat and categorize the attack's type on multi-cloud. The detection and categorization accuracy was 99 % and 93.6 % [33].

ML technology also played an important role in Internet of Things (IoT) security. Novel ML models were designed to cope with security threats automatically. This framework utilized Software Defined Networking (S.D.N.) and Network Function Virtualization (N.F.V.) enablers to reduce various hazards. This A.I. system combined a monitoring agent with an AI-based response agent that used ML-Models to analyze network trends and found aberrations to indicate attacks on IoT devices. The modules applied controlled learning, a disseminated information extraction system, and neural networks [45]. ML-based-IDS security solution was also utilized to prevent threats, network anomalies, and traffic [39]. ML algorithms were also utilized to determine the type of threats. Various supervised and unsupervised and hybrid threat-determining models were formulated for detecting threat type, its purpose (accidental or planned), and motivation (malicious or not) [46, 47]. For developing security procedures in complicated task handling systems, MLSCS-based multi-server management techniques, RL-Q matrix techniques, V.P.C. setups, and V.G.A.N. were suggested for this purpose, which proved beneficial [48]. Another machine learning approach based on K.N.N., AdaBoost, and Naïve Bayes was formulated to explore the various security concerns of the cloud environment owing to access, confidentiality, and authenticity. Improves boosting models worked better than K.N.N. concerning the accuracy and identification time [32]. The feasibility of two renowned ML strategies, ANN and SVM, were investigated to identify intrusion or unusual behavior in the cloud. SVM showed 91 %, and ANN exhibited 92 % malware detection accuracy [49]. Thus, a

large number of studies have been made on cloud security by machine learning in the past, such as SVM based [50, 51], ANN-based studies [52-54], and MLP-based approaches [37, 55]. Decision Tree (D.T.) also proved to be an efficient model for detecting anomalies [38, 56]. ML hybrid models have also been studied in the literature due to their superior performances [40, 57, 58]. D.T. and SVM-based hybrid model was proposed to study intrusion detection [59]. The Bayesian Network-based model also served as helpful for managing the cloud. It served for the detection of attacks [60, 61]. The current state of research on employing ML for CC security in the industry demonstrates its potential for improving threat detection and response, automating security tasks, and enhancing access control and identity management.

Machine learning proves to be a promising approach. However, still, there are challenges to overcome, such as the need for more transparency in how machine learning algorithms make decisions and the need to balance security with usability. There are several research voids in comparing the effectiveness of ML models for cloud security. This research presents an ML-based strategy for securing cloud computing services. Three machine learning models, SVM, X.G.B., and M.L.P., are proposed in this study. Various hypotheses are proposed in this study, which become the basis for the aims and objectives. The summary of the hypotheses is shown in Figure 1.

Hypothesis 1

Machine learning models can be beneficial for determining Industrial cloud cyber security, trust, and privacy issues.

In cybersecurity, machine learning approaches are frequently used to spot vulnerabilities, stop different kinds of assaults, and improve overall security measures. This study addresses cloud security, trust, and privacy issues through a survey from different industrialists, and then ML modules are applied to evaluate this data.

Hypothesis 2

Machine learning technology can help identify cyber security and cloud management gaps.

The gaps in cyber security management are important to determine resolving the issues of the industrial cloud, and ML can be an effective technique for this purpose. Moreover, there are gaps in utilizing ML technology for the cloud. In this regard, data is obtained from the survey regarding the gaps in cloud security and employment of ML for this purpose and evaluated by utilizing SVM, XGBoost, and M.L.P. models.

Hypothesis 3

SVM, XGBoost, and M.L.P. models can be utilized for the detection of threats and can trigger appropriate actions and mitigate them.

To test this hypothesis, these models are designed, and data regarding security threats obtained from the survey is used as input data to run the models. The performance of these models in determining the security threats and triggering appropriate actions is then evaluated using the evaluation matrices.

Hypothesis 4

SVM, XGBoost, and M.L.P. modules' performance vary for cloud computing security.

To test this hypothesis, the performance of these models is compared, and the best-performed method for industrial cloud computing security is determined.

 doi:10.20944/preprints202305.1492.v1

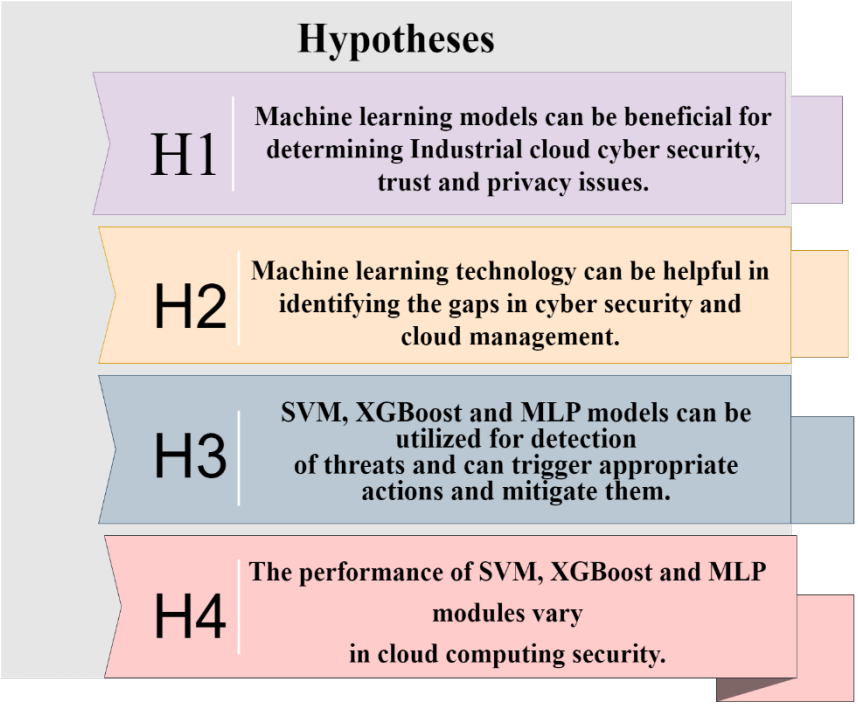

Figure 1. Proposed Hypotheses.

By considering the hypotheses in this study, we set the goals and objectives for this study. The main concern of this study is to formulate a practical strategy for making predictions about the employment of machine learning in the cloud environment. The aims and objectives of the study are as follows.

- Use ML models to sermonize industrial cloud cyber security, trust, and privacy issues.
- Identification of gaps in utilizing the ML approach for cloud security.
- Detection and mitigation of security threats.
- Triggering appropriate security actions.
- Comparison of the performance of SVM, X.G.B., and ANN models in cloud computing security.

There are numerous studies on the ML approach to cloud security. However, as far as we know, studies have yet to make predictions regarding the employment of machine learning and compare the effectiveness of these three models for industrial cloud security. Additionally, it offers experimental findings demonstrating the viability of the suggested strategy. It presents a framework that combines feature selection, feature extraction, and classification techniques for detecting anomalous behavior in cloud data. The research methodology involves surveying industry professionals to understand their perceptions of C.C.S. and the potential of ML in enhancing cloud security. The questionnaire includes 15 questions covering various cloud computing security and machine learning aspects. The survey aims to provide insights into the industry's current security measures, privacy and trust issues, challenges faced, and the potential of machine learning in addressing these challenges. This research concentrates on using ML techniques to improve cloud computing security in the industry. The research explores the current security measures used in the industry and how machine learning can complement these measures to create a more robust and secure cloud computing environment. The paper explores the difficulties in adopting ML for cloud security and suggests solutions.

In conclusion, this paper provides insights into the ability of ML algorithms to boost cloud computing security for industries. This paper highlights the need for continued research and development to create more advanced and efficient security solutions for cloud computing.

Table 1. Literature review summary.

Year	Study	Focus	Key Findings and Limitations
2016	Kaur et al. [32]	Data classification In cloud	Analyze the security issues at authentication and storage. Development of data classification model. Author does not suggest any framework to solve the security concerns.
2017	Salman et al. [33]	Anomaly detection and classification	Detection of attacks and their classification by LR and RF. 99 % detection and 93.6 % classification accuracy by RF. Fail to categorize some attacks.
2018	Marwan et al. [34]	Healthcare cloud data security	Prevent unauthorized asses to healthcare cloud data. Use of SVM and FCM for image pixel classification to ensure security. Only focus on image segmentation for security and privacy and does not mention future challenges.
2019	Subramanian et al. [35]	Cloud cyber security	Avoid static nature for security verification of cloud. Used CNN model for automatic response to threats and save enterprise data. Does not mention type of threats, privacy, trust issues and future challenges in cyber cloud
2020	Praveena et al. [36]	Hybrid cloud security	Reduction of security risks to hybrid cloud by enhanced C4.5 algorithm. Determine the level of security during storing and authorizing the data. Author does not discuss threats and trust issues and future concerns of hybrid cloud.
2020	Wang et al. [37]	DDOS attack detection	MLP based model to detect the DDOS attacks. Detection based on the feature selection and feedback mechanism to for detection error. Model not able to find global optimized feature, feedback mechanism can generate false response.
2020	Chkirbene et al. [27]	Anomalies detection	Classify scheme to protect network from unwanted nodes. Reduce incorrect data issues and differentiate attacks.

			Author does not discuss trust concerns, industrial cyber issues, and insufficient models' comparison.
2021	Haseeb et al. [38]	Health industrial IoT security	Avoid uncertainty in data management of health sector. Data protection by EDM-ML approach and ensures trust between networks. Does not compare performance of Models and not mention future prospects.
2021	Alsharif et al. [39]	IoT security	ML-IDS are used to take account of traffic defects. Offloading heavy tasks from cloud. Does not studied industrial cyber cloud concerns, issues regarding using ML approach for cloud.
2022	Tabassum et al. [33]	QoS security	Neuro-fuzzy approach to study cloud security, reliability, and efficiency. Discuss threats, security, and trust issues. No comparison of model's performance.
2022	Bangui et al. [40]	Threat detection in Vehicular Ad-hoc Network (VANET)	Detection and Prevention of Intrusion in VANET. Use of RF and coresets detection for increasing detection efficacy. Does not provide proper solution to the different types of threats. Lack of performance comparison and trust or privacy factors.

3. Methodology

The methodology includes the data collection, evaluation, and architecture for machine learning, which includes model development, platform selection, data transferring, and cloud platform evaluation. The flowsheet for the methodology is given in Figure 1 below.

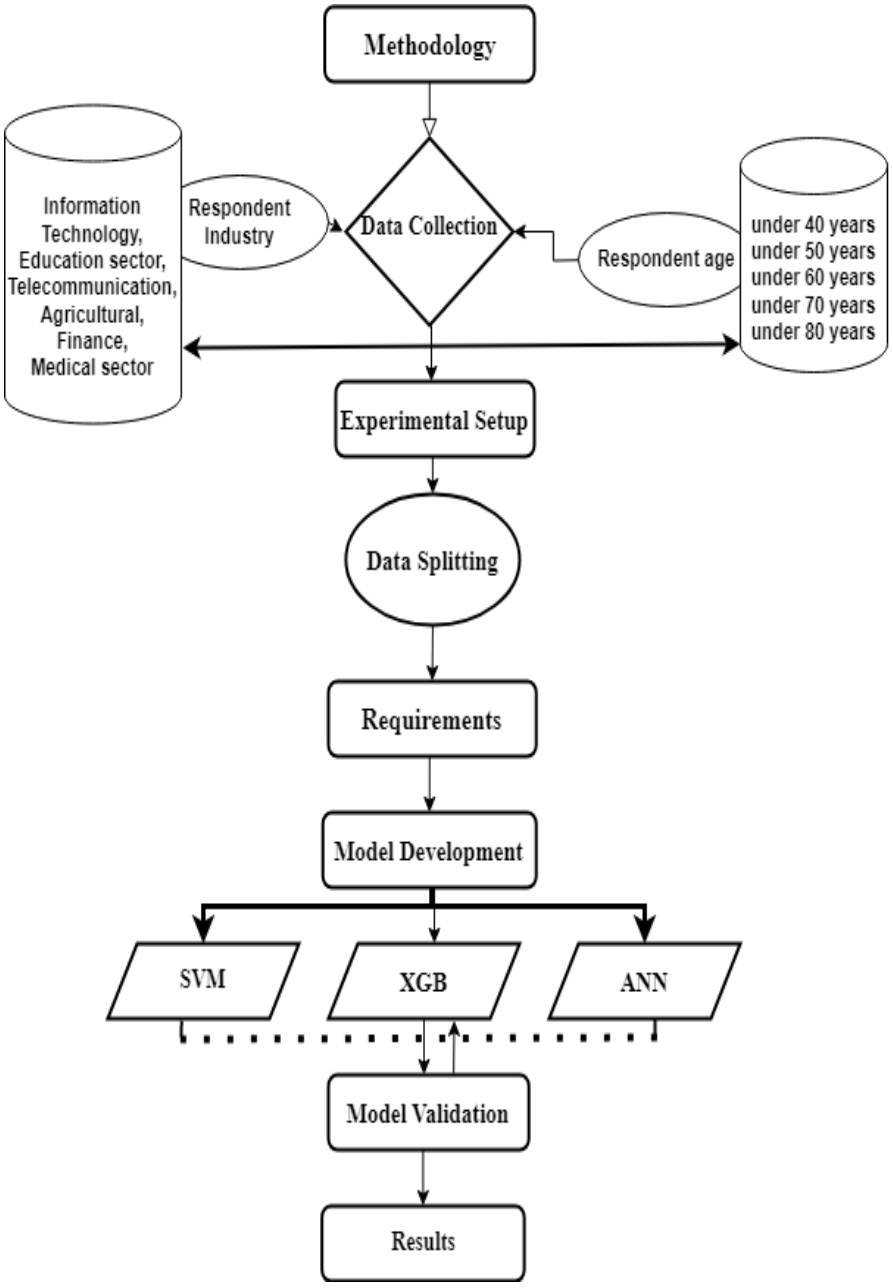

Figure 2. Flowsheet of Applied Method.

3.1. Data Collection

Data is gathered from the survey, which consisted of 15 questions covering every aspect of ML in industrial cloud computing security. The survey is conducted from various industries to collect data regarding the present state of cloud computing security in the industry, the effectiveness of existing security measures, and the potential of machine learning to enhance security. The survey questions are described in Table 2, and Respondent information is also closed in Figure 3. The operational data includes a history of the work completed at each job site. From the operational data, many metrics are derived to reflect a worksite's current status and past events in a predetermined set of variables. The measurements derived from each hour will serve as an input data vector. A worksite's finishing time is represented as the number of hours required before completion; hence the output value of an ML model is the number of working hours required before the worksite is finished.

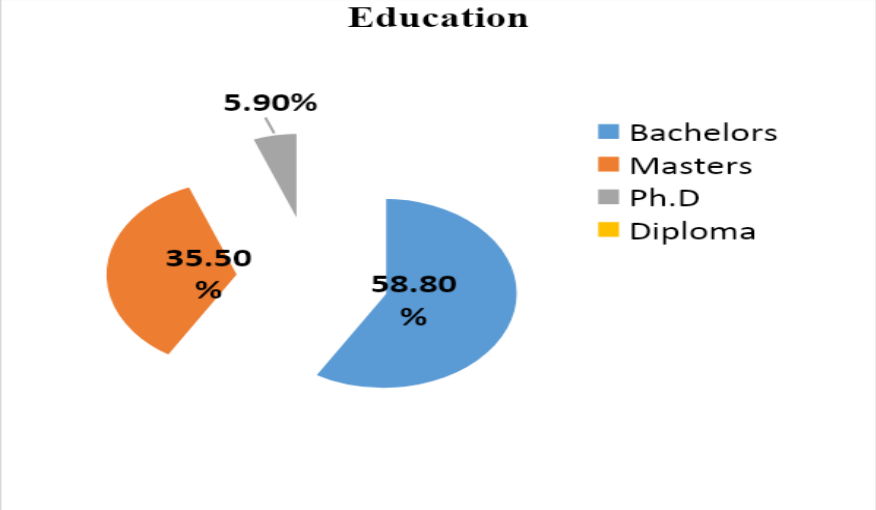

Figure 3. Respondent's Information.

Table 2. Questions for the survey.

Sr No.	Survey Questions
1.	How familiar are you with cloud computing security and machine learning?
2.	Have you or your organization implemented any cloud computing security measures in your operations?
3.	What are the biggest security concerns you have about cloud computing?
4.	How do you think ML can be used to improve cloud computing security?
5.	How confident are you in the effectiveness of current cloud computing security measures?
6.	In your opinion, what are the biggest challenges in implementing effective cloud computing security?
7.	How often do you or your organization conduct security assessments or audits for cloud computing systems?
8.	What role do you think human factors play in cloud computing security?
9.	What measures do you think cloud service providers should take to improve the security of their offerings?
10.	How do you think regulations and compliance requirements affect cloud computing security?
11.	How can organizations ensure their cloud service providers comply with security standards and regulations?
12.	How do you think increasing Internet of Things (IoT) devices affect cloud computing security?
13.	How effective do you believe machine learning has improved the security of your industry's cloud computing operations?
14.	How important is security in your industry's cloud computing operations?
15.	What are your future plans for using machine learning in cloud computing security in your industry?

3.2. Experimental Setup

A cloud-hosted version of the Google Colab platform was employed for this endeavor. Windows 10 and a 2.30 GHz Intel Xeon CPU were used for the experiments. Google collocated with N.V.I.D.I.A. to supply GPUs/CPUs, which were used to speed up calculations. The research was conducted using Python version 3.8, and Google Colab has access to 16 G.B. of RAM for the experiments.

3.3. Data Splitting

 doi:10.20944/preprints202305.1492.v1

Data splitting is a common strategy in ML that divides a dataset into two or more subsets, most commonly a training and testing set. Testing a machine learning model's performance on data it has never seen before is the primary motivation for data splitting. The training set works to train the data while the results or performance of the model is validated on the testing set. This research split the dataset into two parts. 80% of the data is utilized to run the ML models, while 20% is utilized to analyze the ML models.

3.4. Requirements

The first condition for the cloud service architecture is that all machine learning models, including SVM, XGBoost, and Multilayer perceptron, must be implementable. To prevent needing to design architecture on several platforms, all the models must be evaluated on just one platform. The second condition is that the trained algorithm or model must be usable after training for forecasting or prediction.

3.5 Model's Architecture

The data acquired from the survey is the output values for the training data. Since these are continuous, the algorithms utilized in this paper fall within the category of regression and learning. The ML models utilized are introduced in this section.

3.5.1. Support Vector Machine (SVM)

It is the simplest model and requires no hyperparameters. SVM is a linear classifying method that uses projection to transform the initial dataset into higher-dimensional data. When categorizing subsequent data tuples, the hyperplane is established using support vectors and margins to maximize the hyperplane's margins to generate more precise results [62]. Simple linear regression, which has only one input variable, x, and an output variable, y, cannot fulfill the general case where there can be more than one dimension. A multiple linear regression model is utilized, and the equation [63] is given as

$$y_i = \beta_o + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots \beta_k x_{i,k} + \varepsilon \qquad Eq\ 2$$

Where, $\beta_{1\dots k}$ = Regression coefficient
$x_{n,1\dots k}$ = Input values
y = Output or Response
ε = Random residual

3.5.2. Gradient Boosting Model

Gradient boosting models are also used in this paper. The XGBoost technique was used to construct gradient-boosting trees. For XGBoost to function effectively, some parameter adjustment is typically necessary. Although the method fundamentally differs from randomization, boosting models build ensembles from numerous single models. Instead, boosting models progressively add fresh, weak single models to the ensemble (later known as base learners) [64]. Gradient boosting strategies aim to minimize the predicted value of some loss function L(y, F(x)), such as the squared difference between the true output value y and the output provided by model F. (x). K weak base learners ϕ(x) are added to form the model F [65, 66].

$$F(x) = \sum_{k=0}^{k} \beta_k \phi_k(x) \qquad Eq\ 2$$

3.5.3 Artificial Neural Networks

A multilayer perceptron was used to create an artificial neural network. Among the models utilized in this thesis, M.L.P. is the one that needs the greatest fine-tuning[67]. An M.L.P. is a feed-forward system showing that the neurons transmit their outputs forward and do not initiate network cycles [68]. It indicates that a network's output is a predictable function of its input values. M.L.P., which produces a single value from three-dimensional inputs. All input, output, and hidden layers

 doi:10.20944/preprints202305.1492.v1

are closely interconnected. Each hidden layer unit uses all three input layer dimensions as inputs, while all four hidden layer output dimensions are used as inputs by output layer neurons. I and H are the number of input dimensions and hidden layers, respectively. The activation function in output layer σ and hidden layer h, the output is expressed as,

$$y = \sigma\left(\sum_{J=0}^{H} w_{j^2h}\left(\sum_{i=0}^{I} w_{ji^1}\, x_i\right)\right) \qquad Eq\ 3$$

3.6. Evaluation Matrices

This research utilized X.G.B., SVM, and M.L.P. machine learning models. Several measures from the field of evolution. This research used Accuracy, Precision, Recall, F1-score, ROC-AUC, and Confusion matrix, popular machine learning evaluation metrics frequently used in studies to indicate a model's performance [69, 70]. The following equations determine accuracy, Precision, Recall values, and F1 score. The model's effectiveness is studied using a confusion matrix plot. Useful metrics and confusion matrices allow us to assess the total number of correctly and incorrectly labeled classes and precision, recall, accuracy, and F1 score.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad Eq\ 4$$

$$Precision = \frac{TP}{TP+FP} \qquad Eq\ 5$$

$$Recall = \frac{TP}{TP+FP} \qquad \text{Eq } 6$$

$$F1\ Score = \frac{2\ X\ Precision\ X\ Recall}{Precision+Recall} \qquad Eq\ 7$$

4. Results and Discussion:

400 respondents took part in the survey, from which the data was obtained regarding industrial security issues, privacy, and trust issues. Out of all the respondents, most are only graduated and above 30.

4.1. Features Selection

This work aims to establish a practical strategy for making predictions about the employment of ML in the cloud environment. To perform the experiments, this research created a new dataset to implement machine learning models for predictions about the employment of ML in the cloud environment. This study is performed by using selected features. In this scenario, the feature selection approach (Random Forest Feature importance) is utilized to select the most important features for better prediction. The 11most convenient features are used in the experiments. These features are designed to cover this study's objectives, which include identifying gaps and uses of ML models in cloud security, trust, and privacy issues. The visualization of each feature score is shown in Figure 3. This study utilized Extreme Gradient Boosting (X.G.B.), Support Vector Machine (SVM), and Multilayer perceptron (M.L.P.).

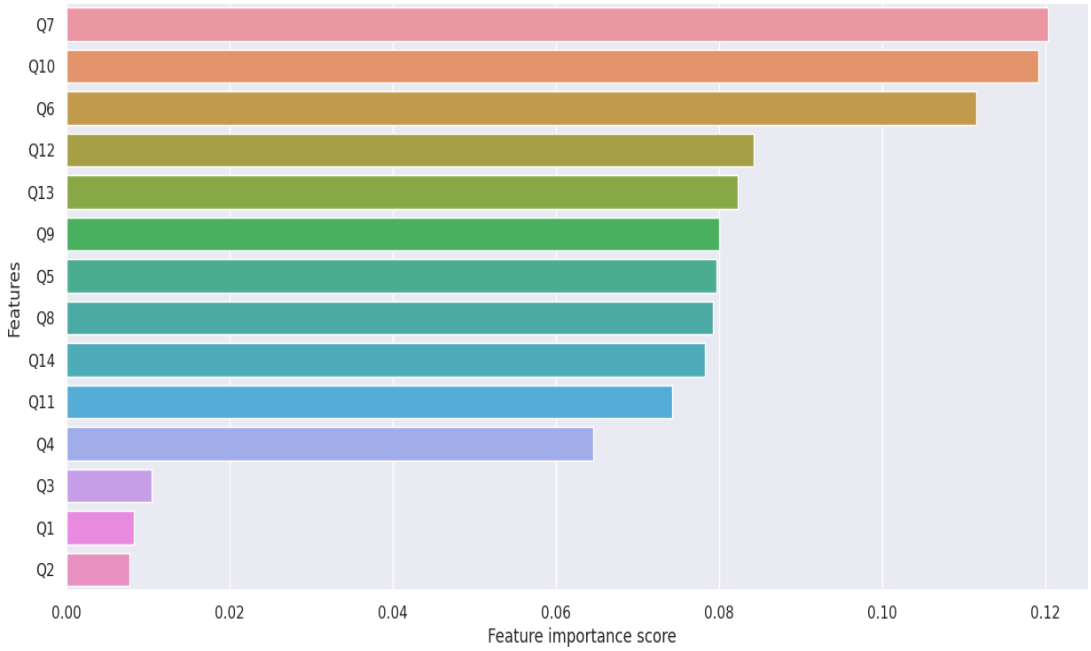

Figure 4. Feature importance score visualization.

4.2. ML Analysis:

This study uses various ML models and algorithms to evaluate the outcomes. The models are assessed by their accuracy, precision, recall values, and F1 score. The ratio of variables in the model properly predicted to the entire dataset is known as accuracy. The model cannot be judged using just this metric. As a result, it is also important to consider the F1 score, precision, and recall parameters. All these parameters are reliable criteria for selecting the optimal model. The performance of the applied models is presented in Table 3.

Table 3. Machine Learning classification results using 11 Feature.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
XGB	97.50	97.60	97.60	97.50	1
SVM	97.35	97.30	97.30	97.30	1
MLP	96.20	96.21	96.20	96.20	99

4.2.1. XGB Model

The X.G.B. model shows the greatest accuracy score of 97.50%. Additionally, it has the highest precision score of 97.60%, meaning that 97.60% of the model's positive predictions are true positives. The model demonstrates a recall score of 97.60%, indicating that out of all the actual positive cases, 97.60% are correctly identified by the model. The F1-score is 97.50%, a harmonic mean of precision and recall. The X.G.B. model achieves a perfect ROC-AUC score of 1, indicating that the model has excellent predictive power. The ROC-AUC curve is plotted and shown in Figure 5.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2023 doi:10.20944/preprints202305.1492.v1

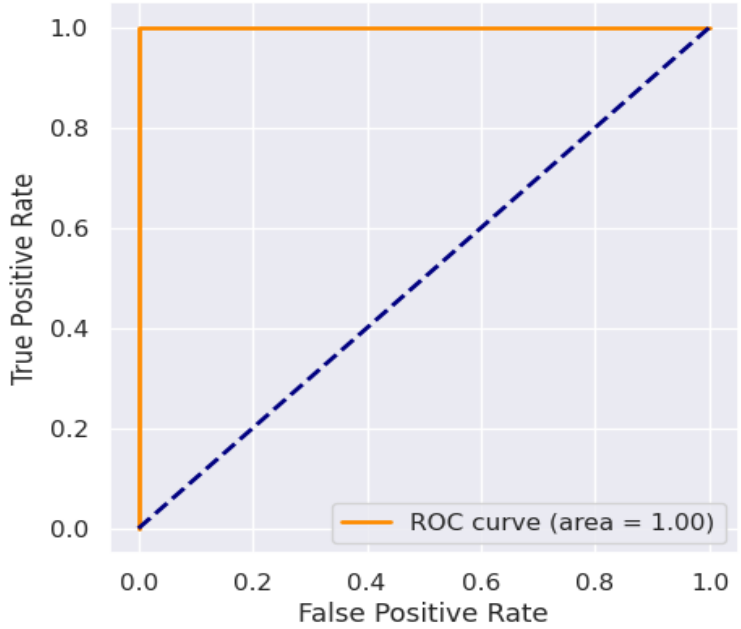

Figure 5. R.O.C. curve visualization of the X.G.B. model.

The confusion matrix in Figure 6 depicts the X.G.B. model's performance on a multi-class classification problem. The problem has four classes (or categories) represented by the matrix's rows and columns. The matrix does not have class labels, but we can infer they are represented by the numbers 1, 2, 3, and 4. The diagonal of the matrix displays the number of occurrences successfully categorized for each class. In class 1, for example, 33 examples are correctly classified as belonging to class 1, whereas one instance is mistakenly classified as belonging to class 2. The off-diagonal elements represent the number of misclassifications between the classes. While one occurrence in class 3 is wrongly identified as belonging to class 1. The confusion matrix demonstrates that the X.G.B. model fared quite well on the classification job, with only a few misclassifications.

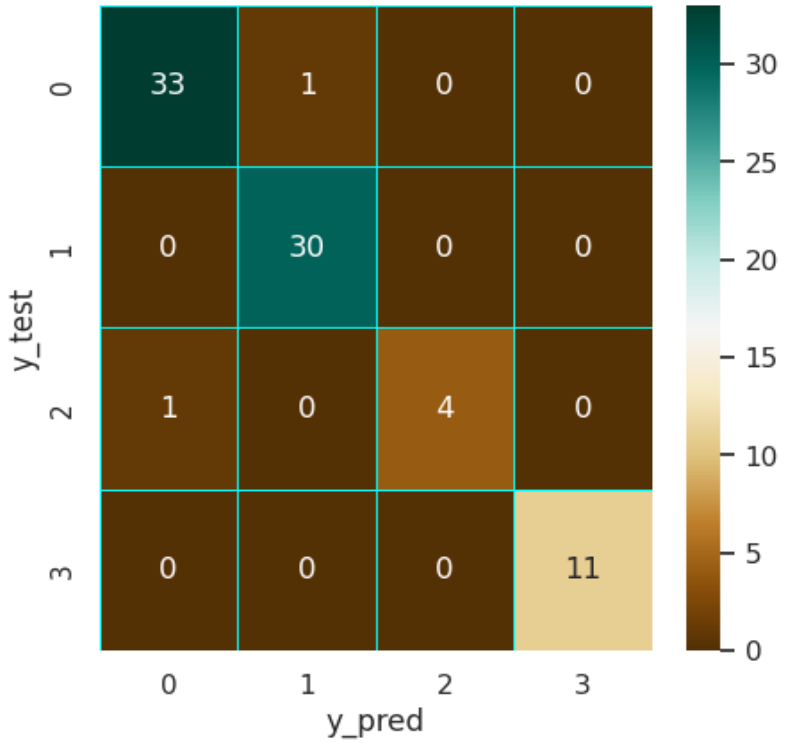

Figure 6. Confusion matrix of X.G.B. model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2023 doi:10.20944/preprints202305.1492.v1

4.2.2. SVM Model

The SVM model depicts an accuracy score of 97.35%. It displays a precision score of 97.30%, suggesting that 97.30% of the model's positive predictions are, in fact, true positives. The model shows a recall score of 97.30%, indicating that out of all the actual positive cases, 97.30% are correctly identified by the model. The F1-score is 97.30%, a harmonic mean of precision and recall. The SVM model achieves a ROC-AUC score of 1, indicating that the model has excellent predictive power. Figure 7 shows the ROC-AUC curve of the SVM model. The confusion matrix of the SVM model is also plotted and shown in Figure 8.

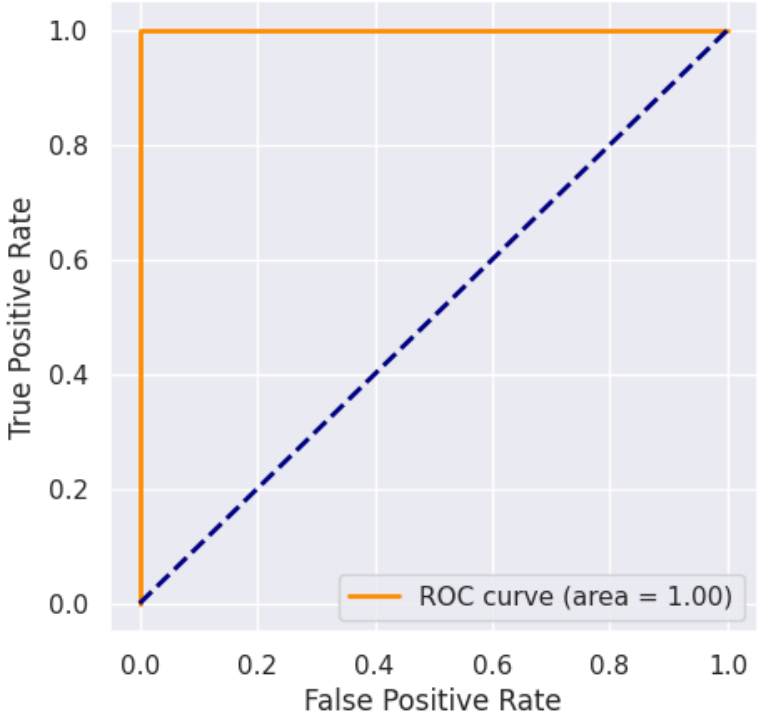

Figure 7. R.O.C. curve visualization of the SVM model.

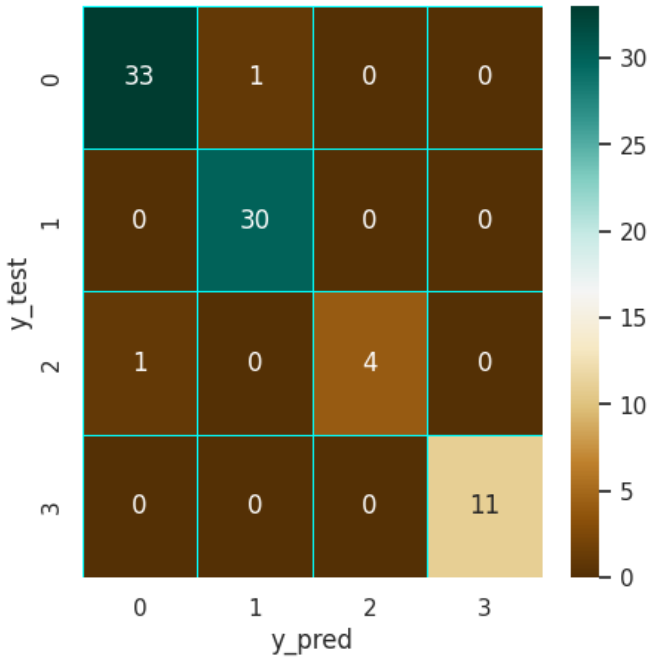

Figure 8. Confusion Matrix of the SVM model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2023 doi:10.20944/preprints202305.1492.v1

4.2.3. MLP Model

The M.L.P. model gives an accuracy score of 96.20%. It shows a precision score of 96.21%, demonstrating that out of all the accurate predictions the model made, 96.20% of them are true positives. The model has a recall score of 96.20%, indicating that out of all the actual positive cases, 96.20% are correctly identified by the model. The F1-score is 96.20%, a harmonic mean of precision and recall. The M.L.P. model has a ROC-AUC score of 99%, as indicated in Figure 9, and the confusion matrix, as indicated in Figure 10, that the model has excellent predictive power.

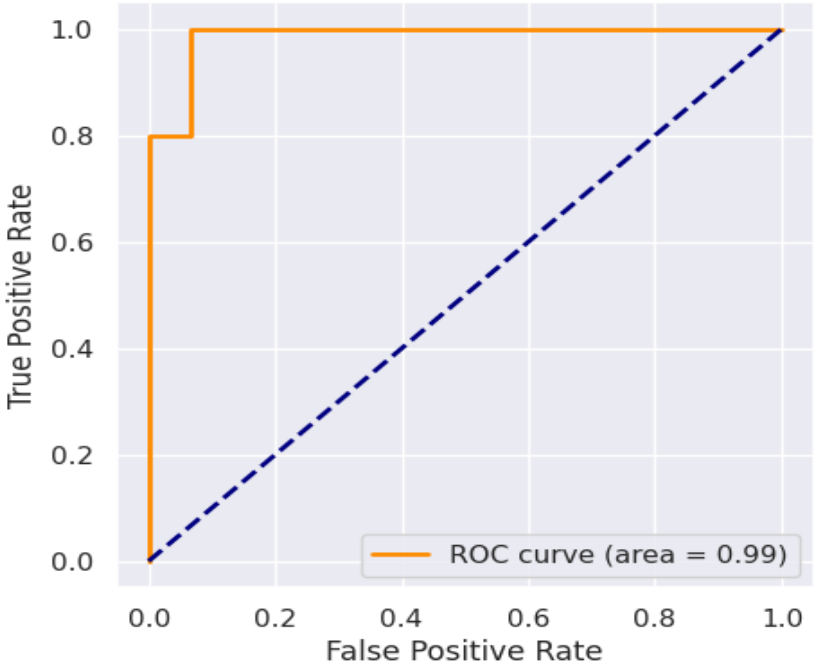

Figure 9. R.O.C. visualization of M.L.P. model.

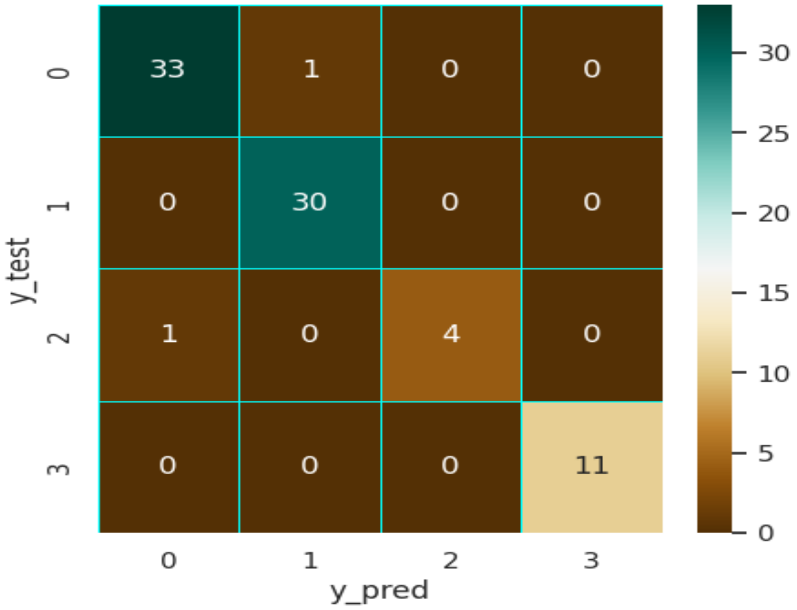

Figure 10. Confusion Matrix of M.L.P. model.

Finally, all the models performed well in accuracy, precision, recall, F1-score, and ROC-AUC score. However, the X.G.B. model achieved the highest accuracy, precision, recall, and F1 score among all the models, while SVM and M.L.P. models also performed well.

The study used ML models to describe industrial cloud cyber security, trust, and privacy issues. The objective was to identify gaps in utilizing the ML approach for cloud security, detect and mitigate security threats, and trigger appropriate security actions. In line with these objectives, the study compared the performance of SVM, X.G.B., and M.L.P. models in cloud computing security. The results showed that all the models performed well in accuracy, precision, recall, F1-score, and ROC-AUC score, indicating their potential to address cloud security issues. Specifically, the X.G.B. model achieved the highest accuracy, precision, recall, and F1 score, which suggests that it can be an effective model for detecting and mitigating security threats in the cloud environment. Therefore, the study's objectives were successfully achieved by demonstrating the potential of ML models in cloud security and identifying the most effective model for this purpose.

5. Conclusion and Suggestion

The employment of machine learning in the cloud environment has been a growing trend in recent years. This research aimed to propose a practical strategy for predicting the employment of ML models in the cloud environment. The study created a new dataset and used a feature selection approach (Random Forest Feature importance) to choose the most important features. The 11 most suitable features were used in the experiments. The selection of these features proved to help determine ML usage in cloud security, gaps, trust issues, privacy issues, and threat detection, which were the main objectives of this study. The study employed three machine learning models, X.G.B., SVM, and M.L.P., for predictions. The experimental results showed that the X.G.B. model outperformed the other models in accuracy, precision, recall, F1-score, and ROC-AUC. The proposed approach of using machine learning models for predicting the employment of ML in the cloud environment can be useful for organizations to make informed decisions. The feature selection approach assists in choosing the most crucial features, improving the ML models' performance. The X.G.B. model showed the highest accuracy and can be used as a reliable prediction model.

Future suggestions could expand this research by adding new features that could enhance the effectiveness of the models. Other ML models, such as Random Forest, Decision Trees, and Naive Bayes, can be utilized to compare their performance with the selected models. Furthermore, the dataset can include more instances and various features. It can assist in creating a machine learning prediction model that is more reliable and accurate for use in a cloud setting. User confidence in cloud computing is increased by achieving robust privacy, and algorithm training efficiency can be increased using sophisticated optimization approaches. These techniques will be effective against serious collusion and dependable but observant servers. Using federated learning, investigators and developers can handle various cloud computing-related problems, such as high communication expenses, confidentiality challenges, statistical variation, and system diversity.

Author Contributions: Z.A..; conceptualization, methodology, validation, Formal Analysis, investigation, Writing- Original draft preparation, S.M.; conceptualization, methodology, writing—review and editing, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5C2A03093690).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: It created a new dataset to implement machine learning models about employing machine learning in the cloud environment.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nassif, A.B., et al., *Machine learning for cloud security: a systematic review.* 2021. **9**: p. 20717-20735.

2. Butt, U.A., et al., *A review of machine learning algorithms for cloud computing security.* 2020. **9**(9): p. 1379.
3. Qayyum, A., et al., *Securing machine learning in the cloud: A systematic review of cloud machine learning security.* 2020. **3**: p. 587139.
4. Bhamare, D., et al. *Feasibility of supervised machine learning for cloud security*. In *2016 International Conference on Information Science and Security (I.C.I.S.S.)*. 2016. IEEE.
5. O'Donovan, P., et al., *A comparison of fog and cloud computing cyber-physical interfaces for Industry 4.0 real-time embedded machine learning engineering applications.* 2019. **110**: p. 12-35.
6. Behl, A. *Emerging security challenges in cloud computing: An insight into cloud security challenges and their mitigation*. In *2011 World Congress on Information and Communication Technologies*. 2011. IEEE.
7. Hassan, W. et al., *Cloud computing survey on services, enhancements, and challenges in the era of machine learning and data science.* 2020. **9**(2): p. 117-139.
8. Kumar, R. and R.J.C.S.R. Goyal, *On cloud security requirements, threats, vulnerabilities, and countermeasures: A survey.* 2019. **33**: p. 1-48.
9. Ige, T. and A. Sikiru. *Implementation of data mining on secure cloud computing over a web API using a supervised machine learning algorithm*. In *Artificial Intelligence Trends in Systems: Proceedings of 11th Computer Science On-line Conference 2022, Vol. 2*. 2022. Springer.
10. Khalid, A., et al., *Understanding vulnerabilities in cyber-physical production systems.* 2022. **35**(6): p. 569-582.
11. Yaacoub, J.-P.A., et al., *Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations.* 2022: p. 1-44.
12. Achar, S.J.I.J.o.C., and S. Engineering, *Cloud Computing Security for Multi-Cloud Service Providers: Controls and Techniques in our Modern Threat Landscape.* 2022. **16**(9): p. 379-384.
13. Vinoth, S., et al., *Application of cloud computing in banking and e-commerce and related security threats.* 2022. **51**: p. 2172-2175.
14. Alsmadi, I., et al., *Vulnerability assessment of industrial systems using Shodan.* 2022. **25**(3): p. 1563-1573.
15. Zahariev, P. et al. *A review of the main characteristics and security vulnerabilities of the wireless communication technologies in the Industry 4.0 domain*. In *2022 Joint International Conference on Digital Arts, Media, and Technology with E.C.T.I. Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & N.C.O.N.)*. 2022. IEEE.
16. Karunasingha, D.S.K.J.I.S., *Root mean square error or mean absolute error? Use their ratio as well.* 2022. **585**: p. 609-629.
17. Xing, J., Z.J.S. Zhang, and C. Networks, *Hierarchical network security measurement and optimal proactive defense in cloud computing environments.* 2022. **2022**.
18. Ukwandu, E., et al., *Cyber-security challenges in the aviation industry: A review of current and future trends.* 2022. **13**(3): p. 146.
19. Takabi, H., et al., *Security and privacy challenges in cloud computing environments.* 2010. **8**(6): p. 24-31.
20. Subramanian, E., L.J.S.O.C. Tamilselvan, and Applications, *A focus on the future cloud: machine learning-based cloud security.* 2019. **13**(3): p. 237-249.
21. Kim, H., et al., *Design of network threat detection and classification based on machine learning on cloud computing.* 2019. **22**: p. 2341-2350.
22. Sharma, V., V. Verma, and A. Sharma. *Detection of DDoS attacks using machine learning in cloud computing*. In *Advanced Informatics for Computing Research: Third International Conference, I.C.A.I.C.R. 2019, Shimla, India, June 15–16, 2019, Revised Selected Papers, Part II 3*. 2019. Springer.
23. Kwabena, O.-A., et al., *Mscryptonet: Multi-scheme privacy-preserving deep learning in cloud computing.* 2019. **7**: p. 29344-29354.
24. Gupta, I., et al., *M.L.P.A.M.: A machine learning and probabilistic analysis based model for preserving security and privacy in a cloud environment.* 2020. **15**(3): p. 4248-4259.
25. Thakkar, A. and R.J.A.o.C.M.i.E. Lohiya, *A review on machine learning and deep learning perspectives of I.D.S. for IoT: recent updates, security issues, and challenges.* 2021. **28**: p. 3211-3243.
26. Kumar, R.S.S., A. Wicker, and M. Swann. *Practical machine learning for cloud intrusion detection: challenges and the way forward*. In *Proceedings of the 10th A.C.M. Workshop on Artificial Intelligence and Security*. 2017.

27. Chkirbene, Z., et al., *Machine learning based cloud computing anomalies detection.* 2020. **34**(6): p. 178-183.
28. Kumar, B., et al., *A Static Machine Learning Based Evaluation Method for Usability and Security Analysis in E-Commerce website.* 2023.
29. Kandi, P., et al. *A Review: Data Security in Cloud Computing Using Machine Learning*. in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*. 2022. IEEE.
30. Mohammad, A.S., M.R.J.C. Pradhan, and E. Engineering, *Machine learning with big data analytics for cloud security.* 2021. **96**: p. 107527.
31. Harmon, R.L. and A. Psaltis, *The future of cloud computing in financial services: A machine learning and artificial intelligence perspective*, in *The Essentials of Machine Learning in Finance and Accounting*. 2021, Routledge. p. 123-138.
32. Kaur, K. and V. Zandu, *A secure data classification model in cloud computing using machine learning approach.* International Journal of Grid and Distributed Computing, 2016. **9**(8): p. 13-22.
33. Salman, T., et al. *Machine learning for anomaly detection and categorization in multi-cloud environments*. in *2017 IEEE 4th international conference on cyber security and cloud computing (CSCloud)*. 2017. IEEE.
34. Marwan, M., A. Kartit, and H. Ouahmane, *Security enhancement in healthcare cloud using machine learning.* Procedia Computer Science, 2018. **127**: p. 388-397.
35. Subramanian, E. and L. Tamilselvan, *A focus on the future cloud: machine learning-based cloud security.* Service Oriented Computing and Applications, 2019. **13**(3): p. 237-249.
36. Praveena, D. and P. Rangarajan, *A machine learning application for reducing the security risks in hybrid cloud networks.* Multimedia Tools and Applications, 2020. **79**: p. 5161-5173.
37. Wang, M., Y. Lu, and J. Qin, *A dynamic MLP-based DDoS attack detection method using feature selection and feedback.* Computers & Security, 2020. **88**: p. 101645.
38. Haseeb, K., et al., *Efficient data uncertainty management for industrial health internet of things using machine learning.* International Journal of Communication Systems, 2021. **34**(16): p. e4948.
39. Alsharif, M. and D.B. Rawat, *Study of machine learning for cloud-assisted iot security as a service.* Sensors, 2021. **21**(4): p. 1034.
40. Bangui, H., M. Ge, and B. Buhnova, *A hybrid machine learning model for intrusion detection in V.A.N.E.T.* Computing, 2022. **104**(3): p. 503-531.
41. Liu, J., et al., *A Bayesian Q-learning game for dependable task offloading against DDoS attacks in sensor edge cloud.* 2020. **8**(9): p. 7546-7561.
42. Yu, Z. et al., *Privacy-preserving federated deep learning for cooperative hierarchical caching in fog computing.* 2021. **9**(22): p. 22246-22255.
43. Mohammad, A.S. and M.R. Pradhan, *Machine learning with big data analytics for cloud security.* Computers & Electrical Engineering, 2021. **96**: p. 107527.
44. Tabassum, N., et al., *Qos-based cloud security evaluation using the neuro-fuzzy model.* Computers, Materials & Continua, 2022. **70**(1): p. 1127-1140.
45. Bagaa, M., et al., *A machine learning security framework for IoT systems.* IEEE Access, 2020. **8**: p. 114066-114077.
46. Masetic, Z., K. Hajdarevic, and N. Dogru. *Cloud computing threats classification model based on the detection feasibility of machine learning algorithms*. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (M.I.P.R.O.)*. 2017. IEEE.
47. Vora, U., et al., *Machine Learning–Based Security in Cloud Database—A Survey.* Machine Learning Techniques and Analytics for Cloud Security, 2021: p. 239-269.
48. Fontaine, J., et al. *Log-based intrusion detection for cloud web applications using machine learning*. In *Advances on P2P, Parallel, Grid, Cloud and Internet Computing: Proceedings of the 14th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2019) 14*. 2020. Springer.
49. Aboueata, N., et al. *Supervised machine learning techniques for efficient network intrusion detection*. In *2019 28th International Conference on Computer Communication and Networks (I.C.C.C.N.)*. 2019. IEEE.
50. Auria, L. and R.A. Moro, *Support vector machines (SVM) as a technique for solvency analysis.* 2008.
51. Xinfeng, Z. and Z. Yan, *Application of support vector machine to reliability analysis of engine systems.* TELKOMNIKA Indonesian Journal of Electrical Engineering, 2013. **11**(7): p. 3552-3560.

52. Michalski, R.S., J.G. Carbonell, and T.M. Mitchell, *Machine learning an artificial intelligence approach*. 1984: Springer.
53. Pandeeswari, N. and G. Kumar, *Anomaly detection system in cloud environment using fuzzy clustering based ANN.* Mobile Networks and Applications, 2016. **21**: p. 494-505.
54. Sheik, S.A. and A.P. Muniyandi, *Secure authentication schemes in cloud computing with a glimpse of artificial neural networks: A review.* Cyber Security and Applications, 2023. **1**: p. 100002.
55. Singh, K.J. and T. De, *MLP-GA based algorithm to detect application layer DDoS attack.* Journal of information security and applications, 2017. **36**: p. 145-153.
56. Ahakonye, L.A.C., et al., *Efficient classification of enciphered SCADA network traffic in a smart factory using decision tree algorithm.* IEEE Access, 2021. **9**: p. 154892-154901.
57. Shon, T. and J. Moon, *A hybrid machine learning approach to network anomaly detection.* Information Sciences, 2007. **177**(18): p. 3799-3821.
58. Rabbani, M., et al., *A hybrid machine learning approach for malicious behavior detection and recognition in cloud computing.* Journal of Network and Computer Applications, 2020. **151**: p. 102507.
59. Peddabachigari, S., A., Abraham, and J. Thomas, *Intrusion detection systems using decision trees and support vector machines.* International Journal of Applied Science and Computations, U.S.A., 2004. **11**(3): p. 118-134.
60. Zimba, A., H. Chen, and Z. Wang, *Bayesian network-based weighted A.P.T. attack paths modeling in cloud computing.* Future Generation Computer Systems, 2019. **96**: p. 525-537.
61. Nie, L., D. Jiang, and Z. Lv, *Modeling network traffic for traffic matrix estimation and anomaly detection based on Bayesian network in cloud computing networks.* Annals of Telecommunications, 2017. **72**: p. 297-305.
62. Meyer, D., F. Leisch, and K. Hornik, *The support vector machine under test.* Neurocomputing, 2003. **55**(1-2): p. 169-186.
63. Jussila, S., *Worksite data analysis using cloud services for machine learning*. 2019.
64. Natekin, A. and A. Knoll, *Gradient boosting machines, a tutorial. 7 (December)*. 2013.
65. Chen, T. and C. Guestrin. *Xgboost: A scalable tree-boosting system*. In *Proceedings of the 22nd A.C.M. signed international conference on knowledge discovery and data mining*. 2016.
66. Nielsen, D.J.N.U.o.S., and Technology, *Tree boosting with XGBoost.* 2016.
67. Bellafqira, R., et al. *Secure multilayer perceptron based on homomorphic encryption*. In *Digital Forensics and Watermarking: 17th International Workshop, I.W.D.W. 2018, Jeju Island, Korea, October 22-24, 2018, Proceedings 17*. 2019. Springer.
68. Guezzaz, A., et al., *A distributed intrusion detection approach based on machine learning techniques for cloud security*, in *Intelligent Systems in Big Data, Semantic Web and Machine Learning*. 2021, Springer. p. 85-94.
69. Rai, A.K. and R.K. Dwivedi. *Fraud detection in credit card data using machine learning techniques*. In *Machine Learning, Image Processing, Network Security and Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30-31, 2020, Proceedings, Part II 2*. 2020. Springer.
70. Jupalle, H., et al., *Automation of human behaviors and its prediction using machine learning.* Microsystem Technologies, 2022. **28**(8): p. 1879-1887.