International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 December 2023* ***Revised:*** *12 January 2024* ***Accepted:*** *20 February 2024*

Analyzing Enterprise Data Protection and Safety Risks in Cloud Computing Using Ensemble Learning

Aiman Shabbir
Department of Computer Science, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan.

Ahmed Selim Anwar
Texas State University

Nazifa Taslima
Sanders college of business and technology University of North Alabama

Md Abu Sayem
Sanders college of business and technology University of North Alabama

Dr. Abdur R. Sikder
Lecturer, San Francisco state university

Gursahildeep Singh Sidhu
Sanders college of business and technology University of North Alabama

Corresponding author: Aiman Shabbir/ aimanshabbir041@gmail.com

Abstract. Nowadays, cloud computing is a significant advancement in the information technology sector. Cloud computing manages and distributes large amounts of data and resources on the internet. In the IT sector, it is used significantly for accessing IT infrastructure via a computer network without necessitating local installations on individual devices. Protecting data security and privacy in cloud computing has become major issue. In this study we used ensemble machine learning algorithms for analysis of cloud computing data, our focus lies analysis of features that effect the data security and privacy threats in cloud computing. Data was gathered through survey online and physical survey method. The data gathering method involved various industries professional's interactions. The survey dataset features consist of security challenges faced by organizations, such as organization size, industry sector, types of data managed, existing security measures, and prevalent security challenges. The primary focus was on evaluating the effectiveness of three machine learning classifiers: Decision Tree, Random Forest, and Support Vector Machine (SVM), which achieved 85.4%, 89.6%, and 88.2%, accuracies of respectively. To enhance predictive accuracy and robustness, an ensemble learning approach using a voting classifier was implemented, resulting in a significantly improved accuracy of 91.5%. The results show that ensemble learning outperforms individual classifiers in predicting cloud data security threats concerns. This paper highlights significant insights for academics and practitioners by implementing ensemble learning approaches that used for significantly strengthen cloud computing security measures, making them more robust to possible attackers.

Keywords: *Cloud Computing, Data Security, Data Privacy, Ensemble Learning, Machine Learning, Security Challenges.*

1. Introduction

Cloud computing is the distribution of computer services such as servers, storage, networking, databases, analytics, and software via the Internet [1]. Because of this technology, customers may access and use resources on a subscription basis, doing away with the need for servicing of hardware and physical facilities. Cloud computing has transformed with its flexibility and scalability, how organizations and consumers manage and use data and apps. Businesses can also benefit from cost savings and scalability, by utilizing cloud computing in many business purposes as they only pay for the resources they use and can easily adjust their capacity based on demand. Because of this, businesses may remain competitive in a market that moves quickly by quickly adjusting to shifting consumer wants and trends [2].

Organizational growth and efficiency are both boosted by cloud computing. It makes it easy to host several people, so they may share resources and work with less effort. Nevertheless, cloud computing is still hindered by security concerns. Many factors have contributed to the problem. The most obvious one is the prevalence of cloud storage among both individuals and businesses. So, keeping the data secure and preventing its loss or alteration as it travels over the network is of utmost importance. Data must be protected in three ways: availability, confidentiality, and integrity. The second definition of unauthorized access is when a malicious actor tries to pose as a genuine user[3].

By relying on cloud services, businesses no longer have to worry about software hosting and server management. In cloud computing, a third party manages and provides access to computer resources (including hardware and software) over the internet. Modern server computer networks and complex software applications enable these services to be delivered[4].

The rapid development of computing, storage, and communication technologies has given rise to a new paradigm in computing known as "the cloud." This model provides users with shared and programmable resources. To set up one or more massive data centers, cloud computing businesses link a plethora of network nodes and devices. Data centers are the focal point of their infrastructure, platform, storage, and software offerings. In terms of recognition, cloud computing is far and away the leader. Cloud storage is an essential client service that is a critical part of cloud computing[5].

Cloud computing is an internet-based service that encompasses a wide array of applications and functionalities. Cloud computing encompasses a wide range of types. Cloud computing is vulnerable to several security threats, such as virtualization, multiple user access, and unauthorized access, all of which pose a threat to the system [6]. The existence of these vulnerabilities poses a risk to the security of sensitive data, thereby undermining the cloud system's reliability. This is due to the potential hazards posed by these defects. Furthermore, the cloud system poses a threat to the confidentiality of sensitive data. As the amount of public data stored in the cloud grows, so does the storage of sensitive information. These facts make it clear that we need to enhance the installation of security measures in cloud computing. When faced with a large amount of data, traditional methods used to identify and address threats, flaws, and undisclosed vulnerabilities may lose their effectiveness. This is due to the exclusion of certain vulnerabilities from the previously provided list. This is due to the use of traditional approaches for problem identification and resolution [7].

Machine learning (ML) methods are increasingly being used to improve data management efficiency and tackle security issues related to cloud platforms. Machine learning is the methodical analysis of algorithms and statistical patterns used by computer systems to implement certain tactics, relying on models and assumptions for assistance. We do this to achieve the desired results. We only use the term "machine learning" to refer to this specific methodology. We can separate various machine learning approaches into three essential categories. The categories are listed below: Furthermore, apart from the supervised, semi-supervised, and unsupervised variations, there is also the alternative of unsupervised learning [8]. Several methods are available for problem resolution. Decision trees, support vector machines, artificial neural networks, and K-means machine learning are all members of this category. Cloud computing security often employs machine learning models, which have experienced a significant increase in popularity in recent years [9].

The revolutionary implications of cloud computing on the online distribution and management of data and resources made the beginning of a crucial point in the history of information technology (IT). Companies can perhaps save time and effort by connecting to their IT infrastructure across a network. Growing interest in cloud computing has created serious privacy and security issues. This is the case even if cloud computing offers some advantages. Because cloud computing systems can handle and store data centrally, cybercriminals target them. This is because potentially sensitive data is really in danger while using cloud computing solutions[10].

By using ensemble machine learning methods, the purpose of this work is to predict data privacy and security risk issues to analyze data related to cloud computing. The data used in this study was collected through survey Procedure, which included find out those people that are from many businesses by using data collection Questionnaire method both online and physical. The survey dataset basically contained following features management kinds, current security processes, company size, industry sector, data management kinds, current security obstacles encountered by these businesses.[11]

1.1. Problem Statement

Infrastructures for cloud computing can be affected by common security issues. However, because more and more businesses are using cloud technology, old security methods aren't working to protect cloud-based systems and data. Since cloud computing models are dynamically flexible, provider-

assisted, and location-transparent, applications and data on the cloud platform are not limited by hardware and safety restrictions. During a security breach, it can be hard to find and separate a specific real item that is a threat or has been hacked. Assets in the cloud are often managed by more than one company, which makes it hard to adopt a unified security plan because different companies have different goals. Also, because the cloud is available and automated resources are shared among many clients, it's possible for people who aren't supposed to have access to see customer data.

2. Literature Review

In this paper, the authors compared the different encryption techniques along with their advantages and drawbacks in terms of protection and safety in the cloud computing environment and compared the advantages and disadvantages of each of them- Cloud-based data safety and protection issues that are discussed in this work are following: Comparison of encryption techniques for data security Data Encryption Standard, Advanced Encryption Standard, Rivest-Shamir-Adleman, and Attribute-based -Comparison of different encryption techniques with advantages and drawbacks. The work presented in this paper is comparison of different encryption techniques used to enhance data security in the cloud[11].

Cloud computing issues develop when inadequate security policies and procedures are applied. Data safety encryption methods are discovered in this study. A holomorphic encryption technique is offered to secure the cloud. A holomorphic encryption technique is offered to secure the cloud. Finding effective encryption methods to provide the highest security of cloud computing[12].

In cloud computing, the major issue is increase the forecasting network capacity using real-time traffic research. The fundamental cause of this problem is the network management plane's ignorance of cloud resource availability, which is only concerned with network resources. Implementing automated fault management and network self-configuration with deep learning (DL) and machine learning (ML) technologies is a viable way to handle these difficulties. Previous research on optical networks has mostly relied on supervised methods for deep learning and machine learning analysis[13].

In this study, the authors investigate machine learning algorithms in the following ways: (i) anomaly detection and predictive analysis and (ii) Adaptive security measures based on evolving threats. (iii)Integration with existing security frameworks for enhanced protection. In this study, the authors highlight the importance for robust data security protocols for effective cloud operation following: (I) Implementation of robust, scalable security measures. (ii) Continuous updates and monitoring to address new threats. In this study, the authors identify potential hazards and issues: (i) data transmission and storage vulnerabilities. (ii) Risks of unauthorized access and data breaches. (iii) Challenges in maintaining compliance with data protection regulations[14].

In this paper, the authors investigate security protocol enhancement procedures. They explore multi-factor authentication. Discusses machine learning application reliability. Evaluates integration with existing IT infrastructure. Discusses cost-effectiveness of security implementations. Discusses consensus on best practices. Explores scope in various cloud scenarios. Examines cross-border data protection. Discusses long-term scalability. Discusses anticipated industry trends[15].

Convolutional neural networks (CNN) are used to detect attacks, use deep learning techniques. Cloud computing offers on-demand services for data processing and storage, including: (i) Cloud computing offers data processing and storage services that may be accessed whenever needed. (ii) It gives the convenience of many data centers located across the globe, where services are available on a pay-as-you-go basis. The security challenges in cloud computing arise due to the system's vulnerability to different security assaults, despite its many benefits. (ii) Common forms of assaults include Man-in-the-Middle, Malware Injection, and Denial-of-Service (DOS). This study operates on the following fundamental steps: (i) Employing machine learning methods to detect and address security issues in cloud systems. (ii) Assessing the efficacy and dependability of various machine learning techniques in assuring cloud computing security. (iii) Enhancing current security mechanisms to enhance the protection of cloud environments[16].

Cloud computing and machine learning applications are as follows: (i) They Facilitate machine learning applications like medical diagnosis, online fraud detection, and email spam filtering. (ii) Collects data from multiple owners for training and classification in the cloud environment. Data Security and Privacy Concerns: (i) Critical hindrances to using machine learning tools due to data security and privacy issues. (ii) Unauthorized entities can detect statistical input data and infer machine learning model parameters. The privacy-preserving model contains the following: (i) It safeguards data without compromising machine learning efficiency. (ii) Uses ε-differential privacy to protect data. Differential Privacy and Fog Nodes: (i) Addresses bandwidth and latency issues. (ii) Adds noise to the owner's site data for protection. Challenges in Cloud Computing :(i) Reviews security conflicts arising from cloud storage. (ii) Emphasizes the need for privacy protection in machine learning tasks.

Analyzes existing solutions and identifies gaps in current cloud security measures. Highlights the need for improved privacy-preserving techniques. Describes the implementation of the PPOD scheme: Integrates differential privacy and fog nodes into the cloud computing framework[17].

Analysis of clustering techniques for anomaly detection**.** Evaluation of dimensionality reduction methods**.** Exploration of feature extraction for enhancing security measures**.** Identification of data patterns without labeled data**.** Application in detecting unusual access patterns**.** Examination of classification algorithms for intrusion detection**.** Assessment of regression models for predicting security breaches**.** Comparison of decision trees, SVM, and neural networks**.** Evaluation of supervised learning for user authentication**.** Role of labeled datasets in improving security measures[18].

This table provides a clear summary of the methodologies, results, and research gaps of previous studies.

Table 1. Existing Stud

Reference	Methodology	Results	Research Gap/Issues Identified
[1]	Description of cloud computing basics	Cloud computing offers services via the Internet, allowing subscription-based resource use	Basic overview; lacks detailed analysis of advanced cloud features and recent innovations
[2]	Analysis of cloud benefits for businesses	Businesses save costs and can scale resources easily, maintaining competitiveness	Limited discussion on sector-specific cloud adoption challenges
[3]	Identification of security concerns	Emphasizes the importance of availability, confidentiality, and integrity of data	Need for enhanced security measures to address unauthorized access and data vulnerabilities
[4]	Overview of cloud service management	Highlights the role of third-party providers in managing computer resources	Insufficient exploration of potential dependency risks and service disruptions
[5]	Examination of cloud infrastructure	Cloud computing relies on massive data centers and shared resources	Lacks detail on the environmental and operational impact of large-scale data centers
[6]	Identification of cloud security vulnerabilities	Lists threats like virtualization and unauthorized access	Limited focus on emerging threats and solutions for multi-tenancy and data integrity
[7]	Discussion on traditional security methods	Traditional methods may fail with large-scale data, highlighting need for improvement	No concrete strategies proposed for modernizing threat detection and mitigation
[8]	Introduction to machine learning in cloud security	ML methods categorized (supervised, semi-supervised, unsupervised) and their applications	Needs more specific examples of successful ML implementations

3. Methodology

We distributed surveys through email and social media platforms to reach a broad audience. We targeted IT managers, data security officers, and cloud service providers to ensure diverse professional insights. We designed questions to explore the types of data managed, current security measures in place, and specific security challenges faced. We made sure to inquire about the data types handled, the security procedures that are already in place, and any particular difficulties encountered with security. We identified and addressed missing data points using techniques such as imputation (filling in missing values) or removal of incomplete records to maintain dataset integrity. We checked for and eliminated any duplicate records to ensure each entry was unique and valid We converted categorical data into numerical formats using techniques such as one-hot encoding or label encoding to enable the data to be processed by machine learning models. We divided the cleaned and preprocessed data into training and testing sets, typically using an 80-20 or 70-30 split, to train and validate the machine learning models. In order to analyze the data, we employed ensemble machine-learning techniques. Up to 85.4% of the time, the decision tree classifier was correct. At 89% accuracy, the Random Forest classifier proved to be rather effective. We found that Random Forests because they

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: 25 December 2023 *Revised:* 12 January 2024 *Accepted:* 20 February 2024

are ensembles, can be computationally demanding and difficult to understand. A remarkable 88.2% accuracy was attained by the SVC classifier. A voting classifier was created by integrating Decision Tree, Random Forest, and SVM models. In order to make the final prediction more solid, we used a majority voting system. The ensemble method is clearly superior to individual classifiers, as our approach attained a substantially higher accuracy of 91.5%.

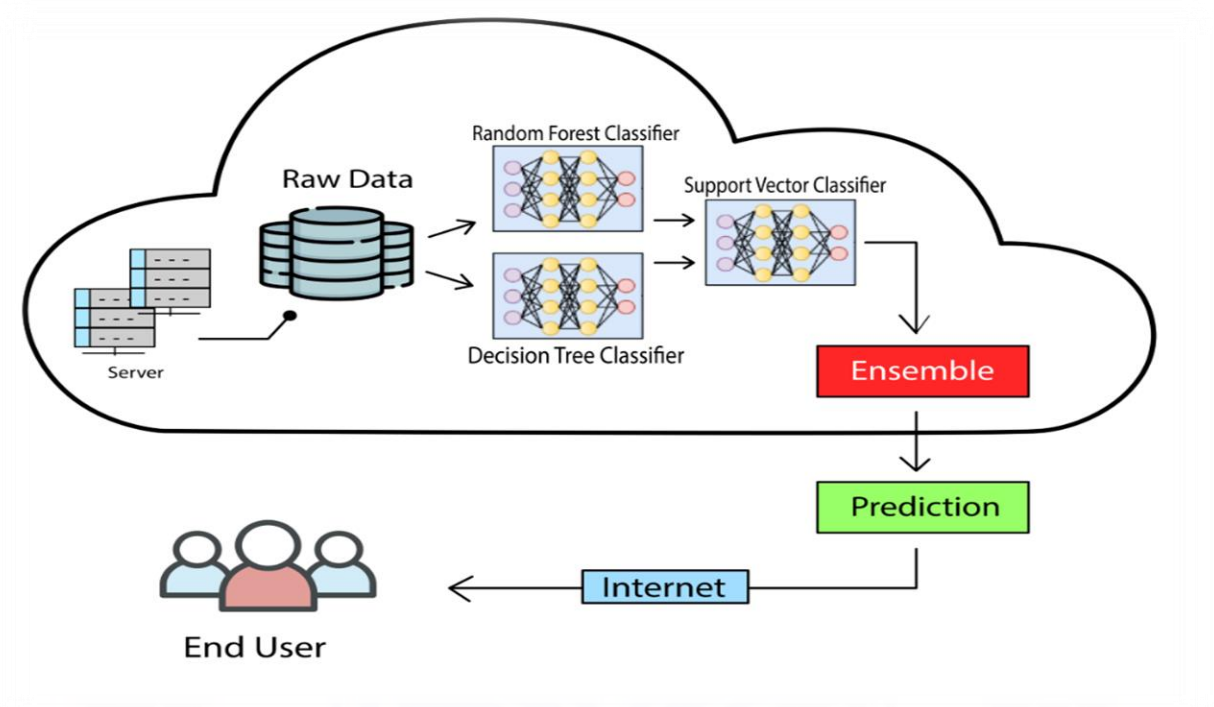

Figure 1. Methodology Framework

The dataset features of our work are mentioned in below table 2. The mentioned below table describe dataset features and dataset features description and features attributes.

Table 2. Dataset Features

Feature	Description	Attributes
Organization Size	Represents the size of the organization, which can affect security resources and threat exposure.	Small, Medium, Large
Industry Sector	Identifies the industry in which the organization operates, each with unique security challenges.	Technology, Healthcare, Finance, Education, Government
Types of Data Managed	Indicates the type of data handled by the organization, influencing data sensitivity and security needs.	Personal Identifiable Information (PII), Financial Data, Intellectual Property, Healthcare Data, Operational Data
Current Security Measures	Describes the existing security protocols implemented by the organization.	Firewalls, Encryption, Intrusion Detection Systems (IDS), Access Controls, Security Audits
Prevalent Security Challenges	Lists the common security issues faced by the organization in the cloud environment.	Data Breaches, Unauthorized Access, Malware Attacks, Data Loss, Compliance Issues

Data Sensitivity Level	Measures the sensitivity of the data handled, which impacts the required level of security.	Low, Medium, High
Cloud Adoption Level	Indicates the extent to which the organization has adopted cloud computing solutions.	None, Partial, Full
Compliance Requirements	Identifies the regulatory and compliance standards that the organization must adhere to.	GDPR, HIPAA, ISO/IEC 27001, PCI DSS, CCPA
Incident Response Time	Measures the average time taken by the organization to respond to security incidents.	Minutes, Hours, Days
Security Budget	Represents the portion of the organization's budget allocated to security measures.	Low, Medium, High
Employee Training	Describes the frequency and extent of security training provided to employees.	None, Annual, Biannual, Quarterly
Third-Party Services	Indicates whether the organization uses third-party services for managing cloud security.	Yes, No
Security Awareness	Measures the general awareness of security practices among employees.	Low, Medium, High
Incident History	Tracks past security incidents faced by the organization.	None, Few (1-2), Several (3-5), Many (6+)
Data Backup Frequency	Describes how often the organization backs up its data, which affects data recovery capabilities.	Daily, Weekly, Monthly, Rarely
Cloud Service Provider	Identifies the cloud service providers used by the organization.	AWS, Azure, Google Cloud, IBM Cloud, Others
Encryption Techniques	Details the encryption methods employed for securing data in the cloud.	None, Basic (AES), Advanced (RSA, Homomorphic), Custom
Authentication Methods	Lists the authentication techniques used to secure access to cloud resources.	Passwords, Multi-Factor Authentication (MFA), Biometric, Token-Based
Access Control Policies	Describes the policies in place to control access to sensitive data and cloud resources.	Role-Based Access Control (RBAC), Attribute-Based Access Control (ABAC), Discretionary Access Control (DAC), Mandatory Access Control (MAC)
Incident Detection Systems	Indicates the presence and type of systems used for detecting security incidents.	None, Signature-Based IDS, Anomaly-Based IDS, Hybrid IDS

Decision Tree Classifier

A Decision Tree is a supervised learning algorithm used for both classification and regression tasks. It operates by splitting the dataset into subsets based on the most significant attribute to predict the target variable. Represents the entire dataset and the first attribute that splits the data. Intermediate nodes representing decisions based on attributes. End nodes representing the final classification or output. Chooses the best attribute to split the data, typically using criteria like Gini impurity or Information Gain. Data is split recursively based on chosen attributes, forming branches of the tree. The process stops when a stopping condition is met, such as maximum tree depth or minimum sample size for a node. The tree structure is intuitive and easy to understand, making it useful for explaining predictions. Can manage both numerical and categorical data without the need for scaling or normalization. Provides insights into the importance of various features in making predictions. Decision Trees help identify the most critical factors affecting data security and privacy in cloud computing. Used to evaluate the classifier's performance, especially for imbalanced datasets. Offers a summary of prediction results to analyze true positives, false positives, true negatives, and false negatives.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 December 2023* ***Revised:*** *12 January 2024* ***Accepted:*** *20 February 2024*

Random Forest

Random Forest is an ensemble learning method that combines multiple decision trees to improve classification accuracy and control over fitting. It is used for both classification and regression tasks, but it is predominantly applied to classification problems. The Random Forest algorithm constructs a multitude of decision trees during training. Each tree is trained on a random subset of the data with replacement (bootstrapping), ensuring diverse trees. During tree construction, each node is split using the best among a random subset of the features, promoting variance among trees for classification, the final output is determined by a majority vote of the individual trees. In cloud computing, Random Forest classify and predict potential security threats based on data.

Support Vector machine

SVC is a supervised learning model that is used for classification tasks. It aims to find the optimal hyper plane that best separates different classes in the feature space. It is a type of Support Vector Machine specifically used for classification purposes. The SVC finds a hyper plane in a multi-dimensional space that separates different classes. The goal is to maximize the margin between the hyper plane and the nearest data points from each class, known as support vectors. Data points that are closest to the hyper plane and influence its position and orientation. The distance between the hyper plane and the nearest data point of each class is maximized to create the best separation. Uses kernel functions to transform non-linearly separable data into a higher-dimensional space where a linear hyper plane can separate the classes. Training SVC can be slow, especially with large datasets and high-dimensional feature spaces. Requires careful tuning of parameters like the regularization parameter (C) and kernel parameters to avoid over fitting or under fitting. The decision boundaries and transformations used by SVC are not as easily interpretable as simpler models like decision trees.

Voting Classifier

A Voting Classifier is an ensemble learning method that combines the predictions from multiple different machine learning models to improve overall classification performance. For classification tasks, it makes decisions based on the majority vote from various individual models. The Voting Classifier integrates the predictions of various machine learning algorithms (e.g., Decision Trees, Random Forests, Support Vector Machines). For classification, each model casts a vote for a specific class. The class with the majority of votes is chosen as the final prediction. Optionally, different weights can be assigned to each model, allowing models with better performance to have a greater influence on the final decision. Allows for both homogeneous (same algorithm with different parameters) and heterogeneous (different algorithms) model combinations. Each model's prediction is considered equally, and the final class is the one with the most votes. Uses the predicted probabilities of each class. The class with the highest sum of probabilities is selected as the final prediction. By leveraging multiple models, the Voting Classifier often achieves higher accuracy than individual models. Combining multiple models helps in mitigating the risk of over fitting that is common with individual classifiers. Can combine different types of models, making it adaptable to various types of datasets and problems. Enhances robustness by averaging out the biases of individual models, leading to better generalization on unseen data. Managing and combining multiple models can increase the computational complexity and difficulty in implementation. The final model's decision-making process becomes less interpretable as it combines multiple models requires careful tuning and selection of the base models and their parameters to achieve optimal performance. Enhances the detection of security threats by integrating predictions from various models trained on different aspects of security data.

$$y\hat{} = \text{argcmax}_{i=1} \sum n\delta(hi(x)=c)$$

n is the total number of classifiers. c represents each possible class label. δ\deltaδ is an indicator function that equals 1. The final prediction y^ is the class that receives the majority of votes from the individual classifiers:

Confusion Matrix

The confusion matrix is organized in a way that allows us to compare the actual target values with the predicted values. Each row of the matrix represents the actual class, while each column represents the predicted class. Here's a generic structure for a confusion matrix for a binary classification problem:

Table 3. Confusion Matrix

Actual / Predicted	**Predicted Positive**		**Predicted Negative**	
Actual Positive	True (TP)	Positive	False (FN)	Negative
Actual Negative	False (FP)	Positive	True (TN)	Negative

True Positive (TP)

The number of instances where the model correctly predicts the positive class.

False Negative (FN)

The number of instances where the model incorrectly predicts the negative class when the actual class is positive.

False Positive (FP)

The number of instances where the model incorrectly predicts the positive class when the actual class is negative.

True Negative (TN)

The number of instances where the model correctly predicts the negative class.

5. Findings

The table 4 summarizes the performance of four machine learning classifiers Decision Tree, Random Forest, Support Vector Machine (SVM), and Voting Classifier predicting data security and privacy threats in cloud computing with each classifier's accuracy, precision, recall, and F1 score. The Voting Classifier, which combines the predictions of the other models, obtained the highest accuracy (91.5%) and precision (91.0%) and recall (91.3%).

Table 4. Results

Classifier	Accuracy	Precision	Recall	F1 Score
Decision Tree	85.4%	84.7%	85.0%	84.8%
Random Forest	89.6%	89.0%	89.5%	89.2%
Support Vector Machine (SVM)	88.2%	87.5%	88.0%	87.6%
Voting Classifier	91.5%	91.0%	91.3%	91.1%

6. Discussion and Conclusions

The Voting Classifier, which combines Decision Tree, Random Forest, and SVM, achieves the highest accuracy of 91.5%. This method utilizes majority voting for the final prediction, resulting in balanced precision and recall scores of 91.0% and 91.3%, respectively. While integrating the complexity of all included models, it remains computationally intensive but offers the most robust performance. The Random Forest classifier outperforms the Decision Tree with an accuracy of 89.6%. Support Vector Machine (SVM) achieves an accuracy of 88.2%, excelling in high-dimensional spaces and offering versatility with different kernel functions. However, it is not ideal for very large datasets and is less interpretable than other methods, with precision and recall scores of 87.5% and 88.0%, respectively. For practitioners, it is recommended to adopt these advanced techniques, continuously monitor and implement industry-specific practices to address unique challenges. Overall, the study highlights how effective ensemble learning is in identifying and addressing cloud computing security issues.

References

1. Raghavendran, C. V., Kavitha, S., & Nandhini, M. (2016). A study on cloud computing services. International Journal of Engineering Research & Technology (IJERT), 4(34), 1-6.
2. Attaran, M. (2017). Cloud computing technology: Leveraging the power of the internet to improve business performance. Journal of International Technology and Information Management, *26*(1), 112-137.
3. Song, D., Dong, J., Qian, X., & Wang, H. (2012). Cloud data protection for the masses. *Computer*, *45*(1), 39-45.
4. Perwej, Y., Khanduja, V., & Juneja, P. (2019). The future of Internet of Things (IoT) and its empowering technology. International Journal of Engineering Science, 2019, 1-10.
5. Li, J., Li, Y. K., Xie, D., & Cai, H. (2018). Secure attribute-based data sharing for resource-limited users in cloud computing. Computers & Security, 72, 1-12.
6. Khorshed, M. T., Ali, A. S., & Wasimi, S. A. (2012). A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. Future Generation Computer Systems, 28(6), 833-851.
7. Shukur, H., Zulkifli, M., Munro, A., & Kazi, T. (2020). Cloud computing virtualization of resources allocation for distributed systems. Journal of Applied Science and Technology Trends, *1*(2), 98-105.
8. Ali, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges. Information Sciences, 305, 357-383.
9. Butt, U. A., Haider, S., & Shah, S. Z. (2020). A review of machine learning algorithms for cloud computing security. Electronics, 9(9), 1379.
10. Zhou, H. (2012). The internet of things in the cloud. CRC Press.
11. Pavani, K., et al. (2023). Data security and privacy issues in cloud environment. In *2023 5th International* Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE.
12. Gottipati, K.N., et al. A Study on Data Security and Privacy Issues in Cloud Computing. in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS). 2023. IEEE.
13. Manimegalai, R., & Durai, U. (2021). Optimizing resource allocation in cloud computing. *Webology*, 18(3).
14. Thabit, F., Alwan, Z., Yahya, S., Al-Janabi, A., & Hashim, F. (2023). Data security techniques in cloud computing based on machine learning algorithms and cryptographic algorithms: Lightweight algorithms and

genetics algorithms. *Concurrency and Computation: Practice and Experience*, 35(21), e7691.

15. Kandi, P., Krishnamoorthy, S., & Somasundaram, K. (2022). A review: Data security in cloud computing using machine learning. In *2022* 5th International Conference on Contemporary Computing and Informatics *(IC3I)*. IEEE.
16. Mishra, J. K., & Janarthanan, M. (2023). Cloud computing security: Machine and deep learning models analysis. In *Macromolecular Symposia*. Wiley Online Library.
17. Gupta, R., & Singh, A. K. (2022). A privacy-preserving outsourced data model in cloud environment. preprint arXiv:2211.13542, 2022.
18. Rani, S., Rani, E., & Kumar, D. (2022). A machine learning approach to analyze cloud computing attacks. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I). IEEE.
19. Manner, J. (2023). A structured literature review approach to define serverless computing and function as a service. In 2023 IEEE 16th International Conference on Cloud Computing (CLOUD) (pp. 516-522). IEEE.
20. Mohammed, C. M., & Zeebaree, S. R. (2021). Sufficient comparison among cloud computing services: IaaS, PaaS, and SaaS: A review. International Journal of Science and Business, 5(2), 17-30.
21. Rashid, Z. N., Zebari, S. R., Sharif, K. H., & Jacksi, K. (2018). Distributed cloud computing and distributed parallel computing: A review. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 167-172). IEEE.
22. Nassif, A. B., Talib, M. A., Nasir, Q., Albadani, H., & Dakalbab, F. M. (2021). Machine learning for cloud security: A systematic review. IEEE Access, 9, 20717-20735.
23. Zeebaree, S. R., Jacksi, K., & Zebari, R. R. (2020). Impact analysis of SYN flood DDoS attack on HAProxy and NLB cluster-based web servers. Indonesian Journal of Electrical Engineering and Computer Science, 19(1), 510-517.
24. Alhenaki, L., Alwatban, A., Alahmri, B., & Alarifi, N. (2019). Security in cloud computing: A survey. International Journal of Computer Science and Information Security (IJCSIS), 17(4), 67-90.
25. Khorshed, M. T., Ali, A. S., & Wasimi, S. A. (2012). A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. Future Generation Computer Systems, 28(6), 833-851
26 Shukur, H., Zeebaree, S., Zebari, R., Zeebaree, D., Ahmed, O., & Salih, A. (2020). Cloud computing virtualization of resources allocation for distributed systems. Journal of Applied Science and Technology Trends, 1(2), 98-105.
27 Ali, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges.Information Sciences, *305*, 357-383.
28 Butt, U. A., et al. (2020). A review of machine learning algorithms for cloud computing security. Electronics, *9*(9),1379.