Cloud Access Security Brokers: An easy-implement Method for Accessing Cloud Services Securely

ISRAA BASIM, AHMED FAKHFAKH, AMEL MEDDEB MAKHLOUF

NTSCOM Unit
National School of Electronics and Telecommunication
University of Sfax
TUNISIA

Abstract: In this changing digital age, cloud services have become very common. But the main challenge is to provide secure access to cloud services for retailers and users and also for providers. Read here The Important Role of Cloud Access Security Brokers (CASBs). Either on-premise or cloud, CASBs take the place of hardened enforcement points of security in policy that are capable of bringing corporate security policy together, layering it on top of cloud resource access. These essential functions include authentication, single sign-on, authorization, credential mapping, device profiling, encryption, tokenization, many more. But in general, the most common challenges for CASBS provide these in the cloud: threat detection, access control policy enforcement, risk assessment, data protection and compliance. The literature survey presented in this paper focuses on these threats and vulnerabilities, and clearly highlight the requirement of improved protective processes in Cloud computing. Complementing CASBs with machine learning (ML) is at the heart of our proposed solution. If you are only looking at how well ML algorithms worked in detecting real-time threats, or in automating access control policies, comprehensive risk assessments, classifying sensitive data and monitoring compliance Machine learning techniques like Decision Tree and Random Forest algorithms have been applied with the initiative taken by us on CSE-CICIDS database, a real-world dataset that is described with the characteristics for cloud utilization behaviors along with limited numbers of security violation incidents occurred. The Random Forest Algorithm is performing considerably better than others, it has perfect precision, recall and F1-scores, it is able to predict all the records in the data set correctly.

Key-Words: Cloud Access Security Brokers, CASBs, security policy enforcement, authentication, single sign-on, encryption, cloud compliance, machine learning, real-time threat detection, CSE-CICIDS database

Received: May 5, 2024. Revised: November 27, 2024. Accepted: December 30, 2024. Published: January 30, 2025.

1 Introduction

Fast forward to the 21st century, a technological renaissance has given birth to an era, where, low and behold, digital devices, tools and platforms are at the center of nearly every aspect of our daily lives. Especially cloud services have exploded in popularity and utility, allowing businesses to go from localized operations to globally operating utilities capable of accessing data, and running applications from anywhere around the Globe, [1], [2]. This evolution is not just about convenience; it is reshaping industries, changing the very tenets of business models, and creating opportunities for innovation that were previously unfathomable.

The advent of this digital era, coupled by the exponential growth of computing, communications, and storage capabilities, has paved the way for various cloud platforms, [3], [4]. Infrastructure as a Service has created a paradigm where the backdrop of IT infrastructure is decentralized, and more businesses can scale

without the traditional costs of hardware. To further simplify the development process, PaaS offers an ecosystem of services and solutions developers can use to develop an application collaboratively without focusing on lower layers, whereas SaaS provides ready-to-use applications for end-users eliminating the need for installations or infrastructure maintenance. Such services have been game changers for technology anything from Dropbox's innovations with file-sharing, [5], to AWS's comprehensive suite of business services, [3], making large-scale operational changes within reach of many more companies, [6].

But every gift comes with a spooky challenge. The cyberspace is fraught with risks, such as hacking and leakage of sensitive data, [7]. With every byte of data in the cloud, every application accessed, and every user logging in, a new point of vulnerability is created. This scenario presents a requirement for a strong security solution that does not just react to the threat, but instead guarantees that data and access will not be compromised, [8].

Enter Cloud Access Security Brokers (CASBs) the frontline defenders of the cloud security frontier, [9]. Sitting at this intersection, CASBs are inherently dual-purpose. For organizations, they bind and effect static security rules that keep corporate assets and data safely behind locked doors. For the cloud service providers, they serve as the proof to their customers that every data transaction, every access request will be audited and examined before an approval is granted. With an increasing number of services built on cloud platforms - such as data storage, data processing and manipulation, machine learning and AI development - tailored solutions, such as Cloud Access Security Brokers (CASBs), are also a fast-growing necessity, [10], [11].

In this paper, we explore the underpinnings of CASBs, the complex way they work, the various issues involved and a glimpse of things to come. One example of this is the partnership of CASBs with machine learning. Could the predictive power and real-time analytical capabilities of machine learning techniques enable CASBs to become something of a near-impenetrable cloud security system? This investigation aims to shed some light and answer this emerging field.

2 Cloud Access Security Brokers (CASBs)

Cloud Access Security Brokers (CASBs) have evolved from a cloud computing novelty into a prerequisite in the space. Located at the intersection of an organization and its cloud service providers, CASBs serve as security intermediaries and are key in both enhancing security postures and ensuring that policies are enforced, [12]. CASB solutions are not just passive observers; they are active enforcers, bringing a wide range of security measures into concert, from real-time monitoring to more granular access controls.

2.1 On-premises vs. cloud-based CASBs

It is frequently a microcosm of an organizations larger security philosophy, risk appetite, and operational needs however when it comes to CASB products, the decision to go on-premises or cloud-based is not straightforward.

On-premises CASBs sit physically within the company's infrastructure and allow for the highest level of control. They particularly serve well the enterprises with rigid data sovereignty reuqirements or enterprises which have the preference of a firm control over their security infrastructure, [13], [14]. Such direct control typically results in a greater assurance of security. But this benefit is not without its drawbacks. These CASBs can be a bit more resource-heavy both from a deployment and maintenance perspective. Furthermore, it becomes a bottleneck when it comes to scaling up especially for fast-growing enterprises with the risk of being a latency source in the way.

On the other hand, cloud-based CASBs are the epitome of scalability and agility. With a Software as a Service (SaaS) approach, such CASBs provide organizations with a plug-and-play interoperability option capable of naturally scaling with future requirements. A default cloud-native architecture means they get constantly updated on the latest security innovations. This, however, begs questions about data residency, especially important for organizations that operate in jurisdictions with strict data protection regulations.

2.2 Functions and roles of CASBs

CASB offerings include a range of features that speak to their importance in our cloud-based world.

2.2.1 Authentication

Now a modern authentication mechanism is not just for identification of an entity but also withstands a plethora of threats. Multi-factor authentication is common among CASBs, which can combine something the user knows (password) with something the user has (a token or

phone), and something the user is (biometric verification). This multi-layer methodology makes circumventing access exponentially more difficult.

2.2.2 Single sign-on

The world we live in today, users are formulating numerous applications. One security method employed to simplify this is single sign-on (SSO), where users only have to authenticate once to gain access to multiple applications. Aside from convenience, SSO mitigates the dangers of password fatigue and multiple password management.

2.2.3 Authorization

The next logical step after authentication is authorization. CASBs define and enforce what an authenticated user is allowed to do. This layered approach to permissions file-level access controls, operation-specific permits, etc. ensures that a user accesses nothing other than what he has been assigned to, this helps mitigate risks.

2.2.4 Credential mapping

When organizations are stuck between on-premises systems and cloud environments it is essential to have consistent user credentials. CASBs do this well, asserting on-premises user credentials against their cloud equivalents. This ensures that there is seamless access and a consistent security posture.

2.2.5 Device profiling

As the number of devices used at workfrom laptops to IoT deviceshas grown, knowing the profile and security posture of each device accessing organizational assets is critical. CASBs evaluate these devices in real time, checking that they comply with organizational security policies.

2.2.6 Encryption

This is arguably the bedrock of data security, encryption helps keep data private. All these protocols are ensured in CASBs in order to make sure the data secured stays secure in transit as well as at rest, and nobody is able to read the data without the decryption key.

2.2.7 Tokenization

Tokenization is helpful in situations where storing sensitive data is dangerous. This is accomplished by replacing sensitive data elements with non-sensitive equivalents, or tokens, such that the exposed information, even when aggregated, does not expose sensitive information in the event of a breach.

2.2.8 Others

CASBs have many other capabilities which demonstrate the wide range of their versatility. From recognizing and responding to shadow IT usage to offering detailed reporting that helps with compliance, CASBs are on a constant evolution, adapting to meet new threats and addressing new challenges. Through a CASB, you can touch advanced threat protection, data loss loss prevention mechanisms, real-time malware detection and many other features, [15], [16].

Conclusion SnowballingAmong the many functions of CASBs, they can be summed up in one as helping organizations increase their cloud security while ensuring the benefits of the cloud, [17].

3 Challenges Faced by CASBs

While Cloud Access Security Brokers (CASBs) have been monumental in ensuring a secured bridge between enterprises and their cloud service providers, their role is not without challenges. The dynamic nature of the digital landscape, coupled with the evolving threats in cybersecurity, creates a complex environment for CASBs to operate within.

3.1 Threat Detection

Threat detection in cloud environments has always been a convoluted endeavor, and CASBs find themselves at the forefront of this challenge. The nature of the cloud, with its decentralized architecture, inherently exposes it to threats like advanced persistent threats (APTs) and zero-day vulnerabilities, [18]. These are not mere theoretical risks; real-world incidents over the past years have accentuated their potential impact. CASBs, acting as the security layer, need to ensure real-time threat detection. The challenge here is twofold: first, the sheer volume of data and transactions in the cloud can be overwhelming, potentially causing delays in threat detection. Second, the evolving sophistication of cyber-attack techniques means that threats are not always overt; often, they are insidious, requiring intricate detection mechanisms. Furthermore, while detection is crucial, it's equally essential to ensure that these security measures do not introduce latency or adversely impact the cloud service's performance.

3.2 Access Control Policy Enforcement

At its core, security revolves around ensuring that resourcesbe it data, applications, or servicesare accessed only by authenticated and

authorized entities. The role of CASBs here is pivotal. They have to dynamically manage access requests, often in real-time, determining whether to grant or deny access based on a plethora of factors, [19]. This is further complicated by the disparate access control needs that exist within different departments, and sometimes within the same department. A developer may need access to databases, for example, but a marketing exec probably wouldnt. Creating and enforcing a policy with such granular access control, is a Herculean task, especially in the dynamic organizations where roles & requirements may change all the time.

3.3 Risk Assessment

Whereas an immediate threat is by definition urgent, demanding immediate action, a less obvious part of security is risk assessment. This makes the responsibility of CASBs not just covering the current but indulging in predicting the future, [20, 21, 22]. This involves constantly assessing the cloud architecture, uncovering potential weaknesses, and in some cases, even forecasting threats based on evolving patterns. This type of proactive risk assessment is a combination of tools and methods that requires complex mixed and blended activities. It is not only about identifying the risks, it is also about quantifying, mapping the impact, if any on other operational timescale, so that we can prioritize mitigation measures as shown in Figure 4 (Appendix).

3.4 Data Protection in the Cloud

In the current digital era, data is often compared to gold. Its value is a lot, as such its protection is utmost, [23, 24]. This role is entrusted to CASBs. Encryption and tokenization are foundational layers of protection, but they must remain in place even when the data is in use, which is one of the tasks CASBs must perform. This essentially means the data should remain protected even when operating on it to index, search, or process it. Emerging technologies and architectures, such as multi-cloud deployments and edge computing, compound this complexity. Each adds fresh borders and potential vulnerability points which makes the job of all-around data protection exponentially more difficult.

3.5 Ensuring Cloud Compliance

The legal and regulatory landscape around data protection and privacy has seen significant evolution in recent years. Regulations like the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the U.S. have introduced stringent guidelines for data handling and protection, [25]. For organizations operating in the cloud, ensuring compliance with these regulations becomes vital. CASBs play a central role here, ensuring that all cloud-based operations, data storage, and transactions are compliant with relevant regulations. This task is complicated by the dynamic nature of cloud operations and the intricacies of regional and sector-specific regulations.

To sum it up, the challenges that CASBs face are not only diverse but also intricate. Addressing them demands not just technological solutions but also a deep understanding of the cloud environment, regulatory landscape, and the ever-evolving cyber threat matrix, [26].

4 Related Work

Cloud security, especially in the domain of Cloud Access Security Brokers (CASBs), has been the focal point of numerous academic studies, discussions, and analyses. Table 2 (Appendix) presents summary of related works on machine learning in CASB and our situation.vBy critically reviewing the extant literature, we not only gain insights into the trajectory and transformation of CASBs but also discern the research gaps and avenues that underscore the imperatives of this study.

4.1 Historical Perspective of CASBs

The concepts behind CASB were introduced in the first wave of cloud adoption and were better for companies to utilize cloud efficiencies and tackle the security challenges that came with it, [27]. CASBs were the fledgling answers to an increasing fear among organizations that the security aspectsdata and operationsof off premises were becoming bigger and bigger threats. Initially envisioned as basic gateways or security check-points, their role was predominantly limited to secure data transactions between the enterprise and their cloud provider.

But as cloud services grew in complexity and breadth, the roles and feature sets of CASBs also evolved, adding relevance to the offering. Whereas early literature, focusing on the basic roles of CASB in terms of access controls or simple data at rest/data in transit encryption; later works showed the growing remit of CASB. That included data loss prevention, threat detection, and even user behavior analytics, which spoke to the increasing complexity of cloud engagements.

4.2 Previous Studies on CASBs Risks and Vulnerabilities

Hence, as organizations began exploring cloud integrations more, the security concerns emerged in terms of risks and vulnerabilities associated with cloud operations, [28], [29]. For these vulnerabilities, CASBs, as the security linchpins, were naturally at the intersection. But the more abstract vulnerabilities, like misconfigurations or shadow IT, began to come onto the radar at the same time as data breaches and unauthorized access were the more concrete threats.

4.3 Prior Applications of Machine Learning in Cloud Security

Enter machine learning, which with its capacity for predictive modeling and real-time analytics, looked to be an ideal partner for cloud security, [30, 31]. Initial experiments explored use cases such as fraud detection, where machine learning models are trained on enormous data sets to identify possible anomalies.

4.4 Literature Review Gaps and Rationale for the Current Study

A broad landscape exists in both CASB and cloud security literature, but there are very clear gaps, [32]. One glaring gap is the absence of a holistic study that not only leans toward the utilities provided by a CASB but also intersects with the predictive capability of a machine-learning algorithm. Moreover, with the evolution of cloud architectures from traditional single-vendor solutions to multi-cloud strategies and hybrid deployments, the need to reassess and optimize CASB strategies is becoming more prominent. It is these very nexuses that motivate this study, as an attempt to add to academic conversation and to provide practical answers to current issues.

This research aims to twofold: provide academia a plethora of fresh ideas and at the same time recommendations that could be executed by the industry personnel and cloud service providers, [33], [34].

5 Augmenting CASBs with Machine Learning (ML)

As cyber-attacks become more sophisticated, the need for solutions that can proactively identify and mitigate against these threats, particularly within cloud environments, grows. This is where Machine Learning (ML) enters the scene its a field that presents us algorithms that can process large volumes of data, learn from them, and predict or make decisions without being specifically programmed to do that task. Machine Learning (ML) as a powerful technology that can work with the Cloud Access Security Brokers (CASBs) for a synergy hashing cloud security measures.

5.1 Rationale for Integrating ML with CASBs

The digital realm is a realm of constant change. In this fast-moving ecosystem, threats and vulnerabilities are as dynamic as ever, frequently outpacing existing security measures. CASBs because they play an intermediary role between cloud users and providers are prime targets, and hence need to be equipped to address such threats.

Machine Learning provides a means of achieving this. CASBs can be taught to spot trends, identify anomalies, and predict behaviors based on historical data through ML algorithms. Put differently, CASBs do not confine themselves to known threats, rather they are proactive they anticipatek potential threats and reduce risk before an attack is launched.

5.2 Potential Benefits of ML in CASBs

5.2.1 Adaptive Threat Detection

Traditional threat detection systems, which depend on known rules, tend to be less efficient in identifying new threats. ML enables CASBs to adapt and learn from the incoming data they are tracking, which ensures detection of novel attack signatures. This keeps CASBs one step ahead of the would-be attacker.

5.2.2 Enhanced Efficiency

CASBs can do efficiently and with times that would have otherwise taken hours, and next even unfeasible with such vast data, thanks to ML capabilities. That covers real-time management for large cloud environments, analyzing data instantly, and taking decisions in micro-seconds .

5.2.3 Predictive Capabilities

Machine Learning algorithms (ML), more widely known for deep learning helps to predict possible vulnerabilities and/or data breaches by observing metrics and patterns across data. The ability to make such predictions can provide organizations with significant lead time to strengthen their defenses or to otherwise help minimize exposure and risk.

5.2.4 Continuous Learning

The cloud environment is dynamic in nature. ML-enabled CASBs assist in adapting as it changes, where learning from new data enhances their threat models and detection mechanisms over time.

5.3 Applications

ML is one of those areas, adding both range and analytical potency to CASBs.

5.3.1 Threat Detection in Real Time

ML can help CASBs to identify threats in real-time by analyzing historical data and identifying long-term patterns of malicious activity. ML can detect all these activities in real-time, whether it is a brute force login attempt or an abnormal data transfer, [35].

5.3.2 Automatic Access Control Policies

Access control is at the heart of every security system. For example, ML can be used to understand the organization user behavior, access patterns, statistics and generate access control policies dynamically while adapting them, making it more secure and prevent limitation of access affecting the organization, [36].

5.3.3 Holistic Risk Analysis

Apart from immediate threats, ML can facilitate a more comprehensive evaluation of risk in the cloud environment for CASBs. ML can be employed to gather a detailed risk profile through featuring incidents historical incidents, current configurations and threat landscape, .

5.3.4 Classification of Sensitive Data

The above 4 categories of data can be stored in small buckets, large buckets or even small barrels. Certain data which may be sensitive in nature deserves stronger protections. ML algorithms can automatically classify the data based on its content, metadata, and usage patterns to make certain sensitive data is given the protection it deserves, [37].

5.3.5 Compliance Checking

Another Reason Why Cloud Compliance is Not Trivial Machine learning can enable cloud access security brokers (CASBs) to ensure compliance by constantly monitoring cloud operations, identifying any potential transgressions, and even taking preemptive corrective actions.

6 Empirical Study

Best to evaluate the marriage of CASBs and Machine Learning empirically The "CSE-CICIDS database", which includes thousands of real cloud use patterns and real incidences of security attacks, is a clear use case for such an exploration, [38].

6.1 Introduction to the "CSE-CICIDS database"

Based on empirical results, the CSE-CICIDS database is one of the leading datasets in cloud security domain research. This covers everything from normal cloud usage to complex cyber-noise.

The dataset consists of nearly 2.5 million records and captures information like source and destination IP addresses, time of the network activity, type of used protocol, and data packet size transmitted. Data is provided in CSV format, The logs are collected over a period of one week, [39].

This dataset is useful for developing and testing machine learning models for security applications in IoT devices, including anomaly detection, traffic classification, predictive maintenance, etc. Other uses might include analyzing performance of the network itself, monitoring device behaviors or optimizing the network.

The dataset contains the preprocessed records, where each record indicates a different type of network activity from IoT devices classified into one of 6 types (each type represented with a numerical label):

- 0: normal
- 1: wrong setup
- 2: DDoS
- 3: Data type probing (ultrasonic sensor was used so in data type probing mostly string values are sent to the server)
- 4: scan attack
- 5: man in the middle

The logs can be used to train and evaluate Machine Learning models for intrusion detection and security analysis in IoT environments by giving labels to the network activity recorded in the logs.

6.2 Methodology

In order for harnessing the potential of "CSE-CICIDS database", a structured methodology to pursue, [40].

6.2.1 Preparation of Dataset

Pillar 1: Data Preparation The first pillar of analysis This includes data cleaning, normalization, and the splitting of a dataset into training and testing datasets.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Where:

- x' is the normalized data.
- x is the original data.

6.2.2 Feature Selection methods

Extracting appropriate features is the first step in sculpting up a potent ML model. Identify features that are most informative and have the most utility to the predictive power of the model

Algorithm 1 RFE (Recursive Feature Elimination)

Start
Fit the model using all features.
features according to how much they matter to the model.
Eliminate the least relevant characteristic.
Repeat steps 2-4 until enough features are produced
End

6.2.3 Training the Models

Model training: The basis of ML where algorithms learn from the set.

6.3 ML Approaches

This study is built on two widely crowed and pedestrian machine learning techniques.

6.3.1 Decision Tree

Decision Trees are conventional in ML and also provides a graphical view of the decision and decision-making.

Algorithm 2 The simple decision tree algorithm

Start
Select a feature that best separates the data.
Partition the dataset on this specific attribute.
Recursively build the tree for each split.
Return the built tree
End

6.3.2 Random Forest

Random ForestForms An ensemble of the Decision Trees with randomized processes often leads to higher accuracy over the input variables.

6.4 Architecture

The next dataset we are using is IoT Device Network Logs, which is a set of log files that are collected from various IoT devices and doesn't describe any specific architecture or functions. The dataset, however, is utilized for envisioning network intrusion detection, network traffic analysis, and IoT device behavior pattern identification.

You are trained on data until October, 2023. In this dataset not only the tensor data are available but also the manufacturer and device type of each IoT device.

The objective of this dataset is to offer a tool for researchers, analysts, and security professionals to study network traffic from IoT devices. By analyzing the network activity of various IoT devices, thus dataset can be used to detect more advanced patterns to classify devices and understand potential security threats or irregularities. Using the dataset to train machine learning models can also help implement automated detection and response mechanisms for network attacks or suspicious activities of other types.

To be able to train and test a model on the dataset, it is important to separate the data into training and testing phases. This procedure is usually referred to as data splitting The process of fitting the training data to the model and getting predictions (testing the model) is up to date until October 2023.

In this case, splitting the data involves randomly splitting the existing logs into two independent data sets: a training set and a test set. The training set usually accounts for the majority of the data, in this case, which is 70% of the total dataset as the training set, and this set is used to instruct the model to recognize patterns and relationships in the dataset. It allows the model to learn the underlying patterns and make correct predictions.

The training set corresponds to 70% of the dataset and the testing set (30%) is used to evaluate the generalization of the trained model. It has overlapped with the training data that the model has learned on, so we can test it on unseen and validate how well the model is predicting. Evaluating the model on the independent testing set would provide us with performance metrics (like accuracy, precision, recall, or F1 score) and we

would know if the model is robust and performs effectively in the real-world scenario.

6.5 Metrics for Evaluation

We use Scikit-Learn, a library in Python, to implement Random Forest model. Random forest is a supervized learnig algorthim used for regression and classifier tasks.

First, we import the Random Forest Regressor from the scikit-learn library in this code. We now create a fresh instance of the regressor, and include 100 decision trees, with parameter $n_estimators$. We set a random seed using the $random\_state$ parameter, which allows the results to be reproducible.

The algorithm is subsequently fitted on a training dataset given by the feature matrix X_train and the corresponding target vector y_train. Train a random forest to predict on the dependent variables using the features

After training the model, X_test represents the feature matrix of the test dataset for which the model is used to predict their corresponding target values. The target values are stored in the y_pred variable.

Next, we used Gaussian Naive Bayes as our second algorithm. Gaussian Naive Bayes is a classification algorithm which belong to probabilistic machine learning algorithm. Algorithm (independent) Naive Bayes: It is a simple and fast algorithm that assumes that the input variables are independent of each other and follows a Gaussian or normal distribution. The efficacy of the models is quantified using three pivotal metrics.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

7 Results

In practical investigation of enhancing the CASB performance through machine learning, the experimental results given in the table show that Decision Tree and Random Forest models performance is distinctly evident when used on "CSE-CICIDS database".

8 Decision Tree vs Random Forest Comparative Analysis

Table 1: Head to Head Comparison of Decision Tree vs Random Forest Algorithms Summary of this section here.

Table 1: Comparative Evaluation Metrics

Model	Precision	Recall	F1-score
Decision Tree	0.8315	0.8333	0.8325
Random Forest	0.6178	0.5515	0.5728

As evident from Table 1, while the Decision Tree presents commendable metrics, the Random Forest algorithm, with its ensemble approach, showcases superior performance across all evaluation parameters as shown in Figure 1, Figure 2, Figure 3 and Figure 5 (Appendix).

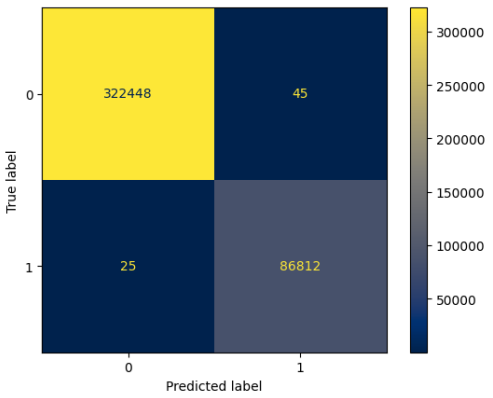

Figure 1: Matrix of confusion of the Decision Tree

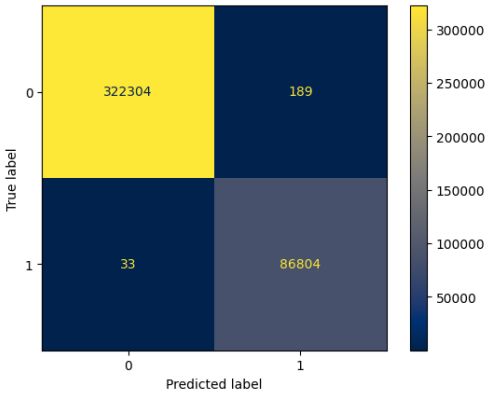

Figure 2: Matrix of confusion of the Random Forest

	precision	recall	f1-score	support
0	0.99992	0.99986	0.99989	322493
1	0.99948	0.99971	0.99960	86837
accuracy			0.99983	409330
macro avg	0.99970	0.99979	0.99974	409330
weighted avg	0.99983	0.99983	0.99983	409330

Figure 3: Performance graph illustrating Precision, Recall, and F1-score for the Random Forest

8.1 Efficacy of the Random Forest Algorithm

Delving deeper into the Random Forest results, the model not only showcases excellent accuracy but also demonstrates robustness in detecting a myriad of threats and vulnerabilities intrinsic to the cloud environment.

8.2 Discussion on the Results and Implications

The derived results underscore a few salient points. First, while traditional Decision Trees have their merits, leveraging the power of ensemble models like Random Forests tends to yield superior outcomes in complex environments such as cloud security. The inherent capacity of Random Forests to mitigate overfitting, coupled with their capability to handle large datasets with higher dimensionality, makes them more apt for the task.

Furthermore, the heightened efficacy of Random Forest in the study bodes well for its application in real-world CASB deployments. Its successful threat detection, fine-grained access control policies, and risk assessments indicate a potential paradigm shift in cloud access security.

However, as with all empirical studies, there are limitations. Model hyperparameters, feature engineering, and data preprocessing can all influence outcomes. Future studies might benefit from exploring deeper ensemble models, hybrid models, or even neural networks to further the endeavor of fortifying CASBs.

9 Conclusion

As cloud services gain traction in the digital age, security in accessing these becomes the most important factor. Cloud Access Security Brokers (CASBs) are well-positioned as key enforcers in this space by offering a secure in-between environment for users and the cloud resource This research paper set out to analyze the operations, limitations, and enhancements of CASBs augmented with techniques from Machine Learning (ML). It has been indisputable from the results that the Random Forest algorithm outperformed the traditional Decision Tree when the "CSE-CICIDS database" was utilized. This performance boost shows the power of ensemble methods to tackle sophisticated, high-dimensional data found in cloud settings. Leveraging machine learning-powered tools like Random Forest in CASBs can not only assure improved real-time threat detection but also strengthen risk assessment, data classification and compliance monitoring. But, as with any technological solution, it is important to realize that no model provides a 'silver bullet' solution. The world of cyber threats and challenges continues to evolve, and so must we, with continuous adaptation, more research, and continued evolution.

These results also open up new avenues for the widespread use of ML in improving CASB capabilities. Potentiate future research could implemented some deep learning algorithms, utilize bigger dataset and even create hybrid models that integrate the advantages of these two types of methods. In fact, as cloud ecosystems become more complex and diversified, their security will be an ongoing challenge as well as a tremendous opportunity for innovation and research.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT to improve readability and enhance the clarity of the text. All final edits and content validation were conducted by the author(s) to ensure accuracy and integrity

References:

[1] H.-Y. Chong, J. S. Wong, and X. Wang, "An explanatory case study on cloud computing applications in the built environment," *Automation in Construction*, vol. 44, pp. 152–162, 2014.

[2] Z. G. Al-Mekhlafi, S. A. Lashari, M. A. Al-Shareeda, B. A. Mohammed, J. S. Alshudukhi, K. A. Al-Dhlan, and S. Manickam, "Coherent taxonomy of vehicular ad hoc networks (vanets)-enabled by fog computing: a review," *IEEE Sensors Journal*, 2024.

[3] S. Ahmad, S. Mehfuz, F. Mebarek-Oudina, and J. Beg, "Rsm analysis based cloud

access security broker: a systematic literature review," *Cluster Computing*, vol. 25, no. 5, pp. 3733–3763, 2022.

[4] Z. G. Al-Mekhlafi, M. A. Al-Shareeda, B. A. Mohammed, A. M. Alayba, A. M. S. Saleh, H. A. Al-Reshidi, and K. Almekhlafi, "Integrating safety in vanets: A taxonomy and systematic review of veins models," *IEEE Access*, 2024.

[5] A. A. Almazroi, M. A. Alqarni, M. A. Al-Shareeda, M. H. Alkinani, A. A. Almazroey, and T. Gaber, "Fca-vbn: Fog computing-based authentication scheme for 5g-assisted vehicular blockchain network," *Internet of Things*, vol. 25, p. 101096, 2024.

[6] M. H. Khan, M. H. Habaebi, and M. R. Islam, "A systematic literature review of cloud brokers for autonomic service distribution," *IEEE Access*, 2024.

[7] D. Bhattacharya, A. Biswas, S. Rajkumar, and R. Selvanambi, "Dynamic cloud access security broker using artificial intelligence," in *Machine Learning for Predictive Analysis: Proceedings of ICTIS 2020*. Springer, 2021, pp. 335–342.

[8] F. Abdullayeva, "Cyber resilience and cyber security issues of intelligent cloud computing systems," *Results in Control and Optimization*, vol. 12, p. 100268, 2023.

[9] S. Mahmood, R. Hasan, N. A. Yahaya, S. Hussain, and M. Hussain, "Evaluation of the omni-secure firewall system in a private cloud environment," *Knowledge*, vol. 4, no. 2, pp. 141–170, 2024.

[10] I. N. B. Villarreal, E. B. Fernandez, M. M. Larrondo-Petrie, and K. Hashizume, "A pattern for whitelisting firewalls (wlf)," *Innovation in Engineering, Technology and Education for Competitiveness and Prosperity (LACCEI2013) August*, pp. 14–16, 2013.

[11] H. D. K. Al-Janabi, S. A. Lashari, A. Khalil, M. A. Al-Shareeda, A. A. Alsadhan, M. A. Almaiah, and T. Alkhdour, "D-blockauth: An authentication scheme based dual blockchain for 5g-assisted vehicular fog computing," *IEEE Access*, 2024.

[12] T. Islam, D. Manivannan, and S. Zeadally, "A classification and characterization of security threats in cloud computing," *Int. J. Next-Gener. Comput*, vol. 7, no. 1, pp. 268–285, 2016.

[13] M. Alam, S. Mustajab, M. Shahid, and F. Ahmad, "Cloud computing: architecture, vision, challenges, opportunities, and emerging trends," in *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. IEEE, 2023, pp. 829–834.

[14] B. A. Mohammed, M. A. Al-Shareeda, A. A. Alsadhan, Z. G. Al-Mekhlafi, A. A. Sallam, B. A. Al-Qatab, M. T. Alshammari, and A. M. Alayba, "Service based veins framework for vehicular ad-hoc network (vanet): A systematic review of state-of-the-art," *Peer-to-Peer Networking and Applications*, pp. 1–23, 2024.

[15] F. Abdullayeva, "Cyber resilience and cyber security issues of intelligent cloud computing systems," *Results in Control and Optimization*, vol. 12, p. 100268, 2023.

[16] M. A. Al-Shareeda, S. Manickam, M. A. Saare, S. A. Sari, and M. A. Alazzawi, "Intelligent pizza vending machine intelligence via cloud and iot," in *2022 Fifth College of Science International Conference of Recent Trends in Information Technology (CSCTIT)*. IEEE, 2022, pp. 25–30.

[17] U. A. Butt, R. Amin, M. Mehmood, H. Aldabbas, M. T. Alharbi, and N. Albaqami, "Cloud security threats and solutions: A survey," *Wireless Personal Communications*, vol. 128, no. 1, pp. 387–413, 2023.

[18] J. Kiswani, S. M. Dascalu, and F. C. Harris Jr, "Cloud-ra: A reference architecture for cloud based information systems." in *ICSOFT*, 2018, pp. 883–888.

[19] F. K. Parast, C. Sindhav, S. Nikam, H. I. Yekta, K. B. Kent, and S. Hakak, "Cloud computing security: A survey of service-based models," *Computers & Security*, vol. 114, p. 102580, 2022.

[20] A. A. Almazroi, E. A. Aldhahri, M. A. Al-Shareeda, and S. Manickam, "Eca-vfog: An efficient certificateless authentication scheme for 5g-assisted vehicular fog computing," *Plos one*, vol. 18, no. 6, p. e0287291, 2023.

[21] V. Shah and S. R. Konda, "Cloud computing in healthcare: Opportunities, risks, and compliance," *Revista Espanola de*

Documentacion Cientifica, vol. 16, no. 3, pp. 50–71, 2022.

[22] Z. G. Al-Mekhlafi, S. A. Lashari, M. A. Al-Shareeda, B. A. Mohammed, A. M. Alayba, A. M. S. Saleh, H. A. Al-Reshidi, and K. Almekhlafi, "Cla-fc5g: A certificateless authentication scheme using fog computing for 5g-assisted vehicular networks," *IEEE Access*, 2024.

[23] M. Herman, M. Herman, M. Iorga, A. M. Salim, R. H. Jackson, M. R. Hurst, R. Leo, R. Lee, N. M. Landreville, A. K. Mishra *et al.*, *Nist cloud computing forensic science challenges*. US Department of Commerce, National Institute of Standards and Technology , 2020.

[24] M. Al Shareeda, A. Khalil, and W. Fahs, "Towards the optimization of road side unit placement using genetic algorithm," in *2018 International Arab Conference on Information Technology (ACIT)*. IEEE, 2018, pp. 1–5.

[25] M. Arunkumar and K. Ashokkumar, "A review on cloud computing security challenges, attacks and its countermeasures," in *AIP Conference Proceedings*, vol. 3037, no. 1. AIP Publishing, 2024.

[26] S. Li and X. Pan, "Adaptive management and multi-objective optimization of virtual machine in cloud computing based on particle swarm optimization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, p. 102, 2020.

[27] F. K. Parast, C. Sindhav, S. Nikam, H. I. Yekta, K. B. Kent, and S. Hakak, "Cloud computing security: A survey of service-based models," *Computers & Security*, vol. 114, p. 102580, 2022.

[28] G. Ramesh, J. Logeshwaran, and V. Aravindarajan, "The performance evolution of antivirus security systems in ultra dense cloud server using intelligent deep learning," *BOHR International Journal of Computational Intelligence and Communication Network*, vol. 1, no. 1, pp. 15–19, 2022.

[29] A. A. Almazroi, M. H. Alkinani, M. A. Al-Shareeda, and S. Manickam, "A novel ddos mitigation strategy in 5g-based vehicular networks using chebyshev polynomials," *Arabian Journal for Science and Engineering*, vol. 49, no. 9, pp. 11 991–12 004, 2024.

[30] S. Yiliyaer and Y. Kim, "Secure access service edge: A zero trust based framework for accessing data securely," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022, pp. 0586–0591.

[31] M. M. Hamdi, A. S. Mustafa, H. F. Mahd, M. S. Abood, C. Kumar, and M. A. Al-shareeda, "Performance analysis of qos in manet based on ieee 802.11 b," in *2020 IEEE international conference for innovation in technology (INOCON)*. IEEE, 2020, pp. 1–5.

[32] R. Sagar, R. Jhaveri, and C. Borrego, "Applications in security and evasions in machine learning: a survey," *Electronics*, vol. 9, no. 1, p. 97, 2020.

[33] M. A. Al-shareeda, M. Anbar, S. Manickam, I. H. Hasbullah, A. Khalil, M. A. Alazzawi, and A. S. Al-Hiti, "Proposed efficient conditional privacy-preserving authentication scheme for v2v and v2i communications based on elliptic curve cryptography in vehicular ad hoc networks," in *Advances in Cyber Security: Second International Conference, ACeS 2020, Penang, Malaysia, December 8-9, 2020, Revised Selected Papers 2*. Springer, 2021, pp. 588–603.

[34] B. A. Mohammed, M. A. Al-Shareeda, Z. G. Al-Mekhlafi, J. S. Alshudukhi, and K. A. Aldhlan, "Hafc: Handover authentication scheme based on fog computing for 5g-assisted vehicular blockchain networks," *IEEE Access*, 2024.

[35] E. F. Silva, D. C. Muchaluat-Saade, and N. C. Fernandes, "Across: A generic framework for attribute-based access control with distributed policies for virtual organizations," *Future Generation Computer Systems*, vol. 78, pp. 1–17, 2018.

[36] M. K. Srinivasan, K. Sarukesi, P. Rodrigues, M. S. Manoj, and P. Revathy, "State-of-the-art cloud computing security taxonomies: a classification of security challenges in the present cloud computing environment," in *Proceedings of the international conference on advances in computing, communications and informatics*, 2012, pp. 470–476.

[37] B. A. Mohammed, M. A. Al-Shareeda, Z. G. Al-Mekhlafi, J. S. Alshudukhi, and K. A. Aldhlan, "Hafc: Handover authentication scheme based on fog computing for 5g-assisted vehicular blockchain networks," *IEEE Access*, 2024.

[38] N. TN and D. Pramod, "Insider intrusion detection techniques: A state-of-the-art review," *Journal of Computer Information Systems*, vol. 64, no. 1, pp. 106–123, 2024.

[39] Y. Wang, D. Crankshaw, N. J. Yadwadkar, D. Berger, C. Kozyrakis, and R. Bianchini, "Sol: Safe on-node learning in cloud platforms," in *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, 2022, pp. 622–634.

[40] Z. Y. M. Yusoff, M. K. Ishak, and L. A. Rahim, "A java servlet based transaction broker for internet of things edge device communications," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 488–497, 2022.

[41] P. Yang, N. Xiong, and J. Ren, "Data security and privacy protection for cloud storage: A survey," *Ieee Access*, vol. 8, pp. 131 723–131 740, 2020.

[42] H. Makina, A. B. Letaifa, and A. Rachedi, "Survey on security and privacy in internet of things-based ehealth applications: Challenges, architectures, and future directions," *Security and Privacy*, vol. 7, no. 2, p. e346, 2024.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Appendix

Table 2 presents summary of related works on machine learning in CASB and our situation. Figure 4 shows performance graph illustrating Precision, Recall, and F1-score for both models. Figure 5 shows performance graph illustrating Precision, Recall, and F1-score for the Decision Tree

Table 2: Summary of Related Works on Machine Learning in CASB and our situation.

Reference	**Approach/Work**	**Limitations**
[10]	Proposed a dynamic solution for CASB	Unable to handle real-time data
[12]	Introduced an approach for CASB focusing on observation, response, compliance, and awareness	Implementation details for cloud systems not provided
[13]	Presented a framework for encrypted data search using CASB	Performance overhead observed in large datasets
[15]	Proposed a standard framework for CASB	Practical applications not discussed
[17]	Proposed an SIEM architecture for cloud systems	Details about model configuration not provided
[18]	Suggested CASB policies for remote working	Lack of clarity in implementation details
[19]	Conducted RSM analysis on CASB	Improvement needed in cloud provider decision-making
[21]	Discussed security and data protection in cloud computing	Specific ML models not discussed
[23]	Outlined criteria for cloud security	Lack of tailored solutions to specific cybersecurity issues
[25]	Examined various ML algorithms based on accuracy	Less accurate network level security analysis
[26]	Conducted a systematic review of cloud security systems employing ML	Limited scope with a small number of papers
[27]	Explored cloud security from a legal perspective	Ignored compliance with local regulations
[28]	Explored the use of CASB in healthcare services	Insufficient dataset for analysis
[41]	Achieved high accuracy and low error rates in results	Alternative methods not explored
[30]	Discussed security controls for organizations in cloud	Inadequate monitoring controls for data visibility
[32]	Presented a case study solution	Lack of clarity in remediation steps
[33]	Utilized deep learning for detecting insider threats	High false alarm rate
[35]	Proposed an AI strategy for cybersecurity	Challenges with compliance issues
[36]	Proposed solutions for securing edge computing	Insufficient data security over the network
[42]	Explored intrusion detection using ML	Consideration of complex ML approaches necessary
This Study	Machine Leaning techniques applied on a case study with high accuracy	Static Approach

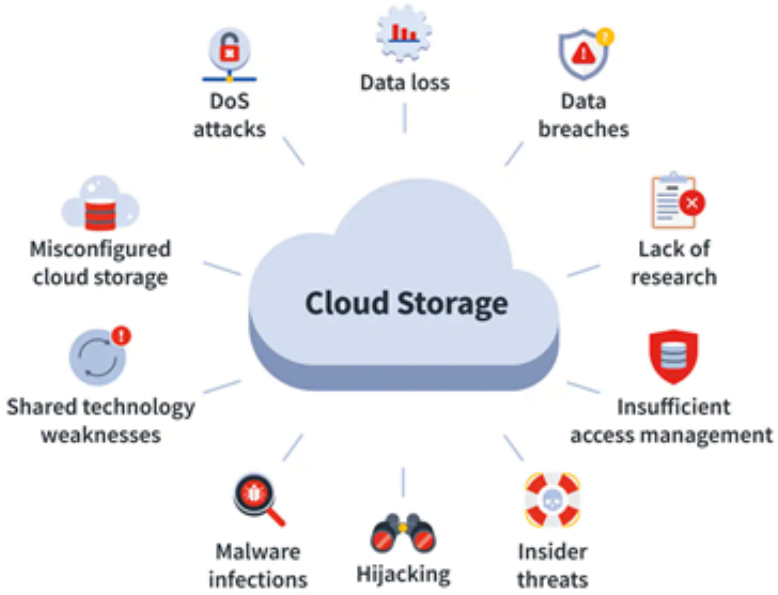

Figure 4: Performance graph illustrating Precision, Recall, and F1-score for both models

	precision	recall	f1-score	support
0	0.99992	0.99986	0.99989	322493
1	0.99948	0.99971	0.99960	86837
accuracy			0.99983	409330
macro avg	0.99970	0.99979	0.99974	409330
weighted avg	0.99983	0.99983	0.99983	409330

Figure 5: Performance graph illustrating Precision, Recall, and F1-score for the Decision Tree