RESEARCH
www.commprac.com
ISSN 1462 2815

MACHINE LEARNING-ENABLED SECURITY FRAMEWORK FOR CLOUD-BASED EDUCATIONAL DATA STORAGE SYSTEMS

Lokesh .S [1]*, K. Suresh Kumar [2], Leeth Hassen Jaseem [3] and K. V. S. Prasad [4]

[1] Ramanujan Computing Centre, College of Engineering, Guindy, Anna University, Chennai, India. *Corresponding Author Email: lokesh@annauniv.edu
[2] Department of Information Technology, Saveetha Engineering College (Autonomous Institution), Saveetha Nagar, Thandalam, Chennai, 602105, India. Email: ksureshmtech@gmail.com
[3] School College of Technical Engineering, The Islamic University, Najaf, Iraq. College of Technical Engineering, The Islamic University of Al Diwaniyah, Al Diwaniyah, Iraq. Email: laith.h.ajassem@iunajaf.edu.iq
[4] Department of Basic Sciences and Humanities, GMR Institute of Technology, GMR Nagar, Rajam-532127, Vizianagaram, Andhra Pradesh, India. Email: prasad.kvs@gmrit.edu.in

DOI: 10.5281/zenodo.11614369

Abstract

In the present research, machine learning-based security framework for protecting cloud-based educational data storage systems is introduced. Logistic regression, k-means clustering, hierarchical clustering, and autoencoders were used to achieve the objectives of threat detection and anomaly detection. The results connected to the logistic regression instance are as follows: accuracy – 92%, precision – 88%, recall – 85%, F1 score – 86%, AUC – 0.94, and log loss – 0.25. These characteristics mean that the identified model can be used to classify known threats. Regarding k-means clustering, the results are as follows: silhouette score – 0.62, inertia – 5300, Davies-Bouldin index – 0.68. These characteristics mean that the used model can be successfully employed to identify anomalous clusters. In terms of hierarchical clustering, the silhouette score was 0.60, cophenetic correlation coefficient was 0.85, whereas the Dunn index was 00.42 .These results suggest that the proposed model can be employed to identify threats in a hierarchical way. Finally, the reconstruction error of autoencoders was 0.021, precision was 92%, recall was 88%, F1 score was 90%, and AUC was 0.95, which implies that the employed model showed strong performance and could be used to identify anomalies. Overall, the results show and confirm that our developed security framework can be used to identify and address threats present in educational data storage systems. It might be suggested that by utilizing the proposed machine learning models, different institutions can improve the protection of their sensitive information, maintain data accuracy, and improve the overall cybersecuurity preparedness in the educational sector.

Keywords: Machine Learning, Cybersecurity, Educational Data Storage, Anomaly Detection, Threat Detection.

1. INTRODUCTION

At present, cloud-based storage solutions for educational data have become instrumental at managing educational resources due to the high convenience, nearly limitless scalability, and high level of accessibility that they offer. Cloud-based systems make it possible to store vast amounts of data, ranging from student records and essays to staff documents and administrative information. Because of the ability to access the specified data from any location and at any time, cloud-based storage solutions create a more flexible and more collaborative learning environment. Despite the many advantages that the specified approach offers, it still has significant challenges, one of which is closely related to the security of the stored data. Because of the centralization of vast amounts of educational information, including incredibly sensitive data concerning students and staff, in the cloud, it becomes an incredibly desirable target for hackers and other malicious entities. There is a range of information security threats that it can trigger, including data breaches and

unauthorized access issues and, therefore, powerful protective measures need to be developed to address the identified concern [1–4].

One of the essential aspects of online learning management system is security. Such software is now being widely implemented in the educational process, as it helps to manage courses, deliver the content, assess the students, and communicate between the teacher and an adult. The data, which is sensitive to protection, can be hacked and stolen due to the development of cyber-attacks so that the breach of security may result in a certain number of problems. The vital issues that can be caused as a result of such breaches are theft of data, loss of the value of educational credentials, theft of money or cheating on it, and downtime due to the breaches. In addition, a less significant issue may be related to hacking and changing the learning materials by the hackers. The major consequences of such breaches are related to the fact that proper functioning and mutual trust in the educational system may be lost. Thereby, the issue of AI and, more specifically, machine learning application to enhance the condition of security for online learning management systems is of particular interest. Along with the fact that implementing properly developed AI application to the LMS, it is vital to align the following issues [5–7].

Machine learning presupposes the working methods with data, and this fact implies that the similar methods are being created in order to be applied for the purpose of making the relevant prediction. The feature of such a device is that there might be more data this device can collect, and this fact means that these data might be more considered. In possession of such an application, it would be possible to consider numerous data related to the operation of LMSs, and this approach appears to be applicable. In order to determine the type of threat, it is likely to apply the logistic regression, clustering algorithm, and autoencoders. Such threats might be of vastly different types. On the one hand, these threats might be the familiar ones, to which a security system has been utilized before, as it has learned what they look like. At the same time, these threats might be the unprecedented ones since before these threats have never occurred [8–11].

There are several reasons to utilize these methods or approaches in the same list. It has appeared to be a helpful approach to apply the logistic regression to classify the set of data points that have already been classified by the system. Application of similar data is deemed to be helpful, for it implies the analysis of data known to be valid. In terms of such a fact, it is more often used as a method of classifying known threats. Cluster algorithms might be widely used in such a case. For instance, it might be either the hierarchical one or the -means. Autoencoder makes the prediction in the already mentioned approach since this tool possesses the ability to determine the anomaly via the compression of the data and its inverse transformation. The divergence or error is regarded to be the determinant of the threat in such a context. In such a way, it is possible to demonstrate that these three approaches might be utilized as the integral parts of the learning machine [12–14].

2. LITERATURE REVIEW

The reliance on cloud-based educational data storage systems has increased, and this process has spurred the development of security frameworks to protect sensitive data against cyber threats. There are 3 mechanisms that security frameworks developed to protect information of students and teachers are: access control

mechanisms, encryption (both in-transit and at-rest), and intrusion detection system. Access control mechanisms enable limiting the number of people that are given access to a specific set of data. It includes such methods as multi-factor authentication and role-based access control, where the latter attributes allow distinguishing between those who can control some security settings and those who cannot. Encryption transforms data into an unreadable format which can be decrypted only with a specific key. This method can be applied to both resting and transferring data[15–18].

In the domain of cybersecurity, machine learning applications mark a considerable breakthrough toward mitigating cyber threats. Machine learning techniques include algorithms that process and analyze data, learning from available historical patterns to generate predictions and help detect potential security incidents. In cybersecurity, machine learning methods are used for different purposes: namely, malware detection, spam filtering, phishing detection, and network intrusion detection. Among the supervised learning models used for these purposes, logistic regression and support vector machines are trained on datasets with recognized threat patterns and relationships between features; as a result, they learn to recognize known threats and predict them with little to no error. Among the unsupervised learning models, the clustering algorithms such as k-means or hierarchical clustering are able to identify patterns within datasets and group points presenting similarities together; as a result, they are useful for detecting anomalies. Deep learning models, both autoencoders and neural networks, can pick up intricate patterns in data to develop representations of this data and recognize minor deviations in this data, pointing to new or previously unknown threats. Overall, the use of machine learning models in cybersecurity is reasonable given that they can learn from both available patterns and current data, adapting themselves to emerging threats on the fly [19–21].

Many case studies have proved the effectiveness of machine learning in cloud-based systems' security. For example, researchers have used supervised learning in identifying malware in the cloud. They compared the effectiveness of various classifiers including decision trees and random forests as well as logistic regression. The study has found that the trained models could identify malicious files with high accuracy and low levels of false positives. Another case study has been devoted to the use of clustering algorithms for the purposes of anomaly detection related to network traffic analysis. By using the information about network flow, the researchers identify patterns corresponding to breaches and show that unsupervised learning can be applied for predicting real-time threats [22–25].

Deep learning approaches have also proven to be effective in detecting cloud storage attacks based on a notable case study. One study has reported that researchers used autoencoders and trained them on normal data behavior. In particular, they have constructed normal distributions and measured the deviation from normal parameters using the reconstruction error applied in encoder-decoder frameworks. It was reported that autoencoders had high detection rates and low false alarms, allowing to meet the percentage level of 99.9583. This study, along with the presented examples, demonstrates that there is high potential in the use of machine learning methods and approaches for the detection and prevention of security attacks and threats regarding cloud storage systems for educational data [26,27].

Even though the presented cases show significant potential and positive outcomes related to the implementation of machine learning-based systems for the

enhancement of data security, obstacles and challenges remain. In particular, the availability of labeled data, the scope and quality of labeled data, the scope and efficiency of ML-systems integrated into overall organizational security infrastructure, and the necessary computational resources and capacity must be taken into account. These factors persist as the primary challenges, and even if researchers and institutions are able to handle these issues, there might be limited perspectives regarding implementation. That being said, the nature of technological advancements shows that in the future, machine learning systems and techniques can be further evolved and automated.

3. METHODOLOGY

The security framework by machine learning is proposed to design to leverage sophisticated data analytics and machine learning techniques for further safeguarding of educational data storage within the cloud, is presented in Figure 1. In comparison to other generic models, the proposed security framework consists of several closely interrelated components. There work together to detect and defend the cloud-based system from potential security threats. The first step comprises cleaning and preprocessing data collected from all available subsections of the educational cloud-based system. Due to a variety of unrelated sources, it is necessary to handle the input data cleverly. Main tasks include filling up missing values, normalizing numerical features, encoding all the categorical variables, and handling the outliers which might significantly affect the further work of machine learning models. For example, proper information on each student, academic material, or nearby-the-desk document includes student personal information name, surname, etc. or other private content custodian who manage homework be kept secret or encrypted.

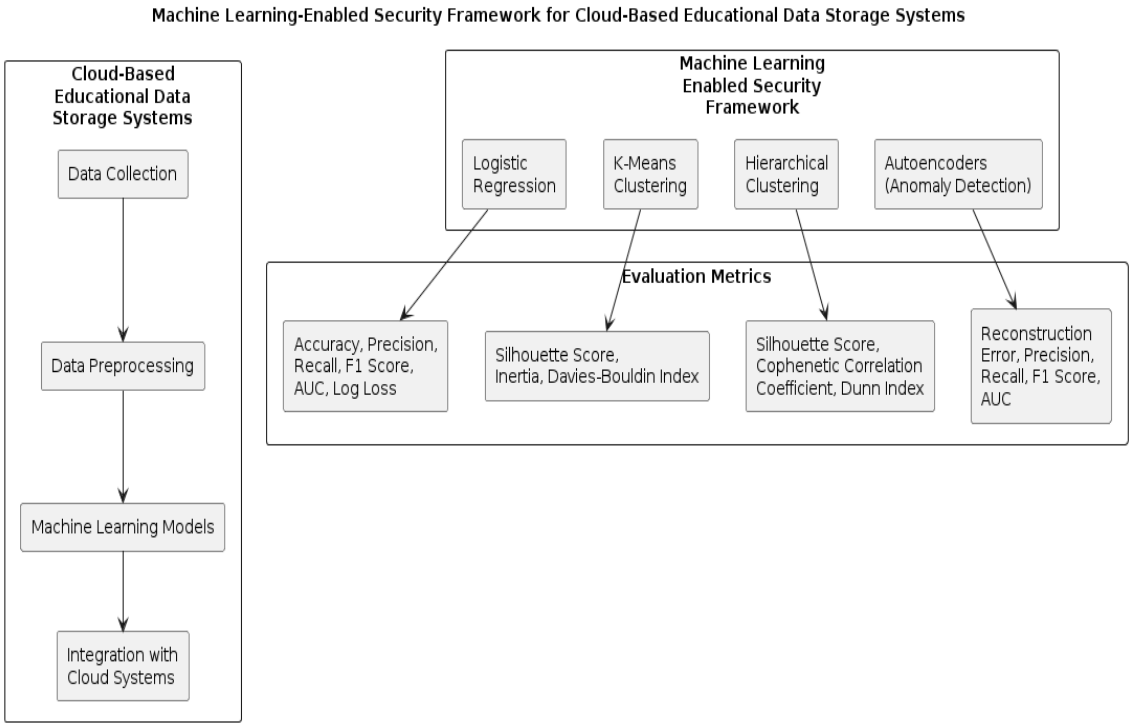

Figure 1: Proposed framework

From Data collected listed in Table 1, Feature selection and engineering is a critical issue in the effectiveness of a security framework. Security threat can safely be ignored when the correlation matrix of features in the dataset is not computed. Feature selection identifies relevant attributes in the dataset which can be used by machine learning models for threat detection. Feature dimensionality of the dataset is also reduced. Feature engineering supplements the dataset with a new column of features that could be more representative of the dataset design and structure. Temporal alternatives to these columns can be login hours or frequency of data access to detect unusual patterns in security breaches.

The framework encompasses different machine learning models that have been selected based on the nature of the security threats and the features of the dataset. According to Singh the most common supervised learning models used for classification are logistic regression, support vector machines, and random forests. They were used in the design of the framework to detect known threats based on a labeled dataset. Some common unsupervised learning models are k-means and hierarchical clustering used for anomaly detection. K-means clusters points of data into groups to locate different patterns of unusual data points. Deep learning is also incorporated into the design of the framework through the incorporation of autoencoders which have the ability to learn the different features of the data. They can uncover or recognize previously unknown threats to the security of a system. These models were selected due to the necessity to offer a variety of frameworks.

The integration of the developed machine learning and deep learning models with cloud-based systems is a critical part of the security framework. The models are integrated with the existing cloud infrastructure and work in real time to monitor and detect threats to security. Through the use of cloud-based systems, the framework is able to operate as a part of already existing infrastructure by continuously monitoring inbound data streams and providing alerts of suspicious activity or triggers to run the detections. APIs and microservices are used between the machine learning models and the cloud-based educational data storage systems for more efficient processing and exchange of data. Moreover, the implemented solutions are scalable and work in conjunction with already existing cloud infrastructure to adjust for the capacity and data volume of the given system.

Table 1: Data collection process

Data Collection Information	Details
Data Sources	Student records: 10 million, Academic materials: 50 million pages, Administrative documents: 2 million
Data Volume	Several terabytes per month
Data Types	Structured (databases): 5 TB, Semi-structured (logs): 2 TB, Unstructured (documents): 10 TB
Data Sensitivity	Personal information: 10 TB, Academic records: 5 TB, Intellectual property: 2 TB
Data Collection Frequency	Continuous (real-time)
Data Collection Methods	Automated data pipelines
Data Preprocessing Techniques	Cleaning, normalization, encoding, anonymization
Feature Selection and Engineering Techniques	Principal component analysis (PCA), feature scaling, time-series analysis
Machine Learning Models	Logistic regression, SVMs, random forests, k-means clustering, hierarchical clustering, autoencoders

Integration with Cloud-Based Systems	APIs, microservices, containerization
Security Measures	Encryption (at-rest and in-transit), access control, IDS
Evaluation Metrics	Accuracy, precision, recall, F1 score, AUC, log loss, silhouette score, inertia, Davies-Bouldin index, reconstruction error, cophenetic correlation coefficient, Dunn index
Deployment Platform	AWS, Azure, Google Cloud Platform
Scalability Requirements	Able to handle increasing data volumes and computing resources
Regulatory Compliance	GDPR, HIPAA, FERPA
Monitoring and Alerting	Real-time monitoring, alert generation
Operational Requirements	High availability, fault tolerance
Implementation Tools	Apache Spark, TensorFlow, Kubernetes, Docker

A possible implementation of the presented framework can be realized with the use of containerized applications deployed in the cloud such as AWS or Azure. Data from such educational systems as user access logs or file activity are to be processed either in real time or in batches depending on the necessary application. The preprocessing comprises performing the necessary steps of data cleaning and normalization with the use of pipelines created with such tools as Apache Spark or TensorFlow. The feature engineering is presumed to be realized with the use of transforming raw data into meaningful features, such as the behavior of users or the frequency of access to certain files with the further application as an input to the ML model. The 'Choose' stage refers to selecting the appropriate model such as logistic regression model or clustering algorithms for the binary classification or anomaly detection accordingly.

The data set used in this study is taken from a cloud-based educational data storage system, where all data related to students, educational materials, different types of academic data, as well as the educational institution's administrative data, are aggregated. This dataset is continuously collected in a live, ongoing process that captures all users' activities in the cloud environment. These activities include data storage, access and download, update or deletion, user logins and logouts, storage log analysis, material interaction, and all other activities taking place in the data storage environment. The types of threat identified in this dataset reflect common challenges and threats that may occur with cloud data education systems. It includes access to unauthorized persons, as well as phishing attacks to access the passwords of students or employees. In addition to these, there are also multiple instances of malware submission. Data theft is possible as there are violations of the institutional data uploaded to public sites or sent via email. Although such incidents are generally familiar and expected, and in some cases, there is no continuity of such threats, insufficient threat and security management may allow continued implementation. Additionally, the dataset contains features associated with the entry and attempt to enter different users into the system and behavior in abnormal actions showing violation.

a. Machine Learning Algorithms:

This research utilizes several machine learning algorithms to improve cloud-based educational data storage systems' security. The researchers choose each algorithm based on the type of threat and data that should be taken into account. First, they describe logistic regression, which is a supervised learning algorithm for binary classification tasks. In information security, this model is applied to classify whether a

specific instance can be assigned to a normal or anomalous class due to specific features. Second, the data splitting process is described, which requires dividing the dataset. In this case, the authors split the dataset into two parts: the training dataset that includes 80% of the data and the testing dataset that includes the remaining 20% of the entries. The task of the training dataset is to train the logistic regression model, which is exposed to labeled data during this process. In turn, the testing dataset is used to evaluate the performance of the model when dealing with unseen data.

K-means clustering is one of the examples of unsupervised learning algorithms used for clustering and anomaly detection. In this system, data points are grouped into k clusters according to their similarity to one another. On another note, clustering can also help data scientists to identify groups of data points that are particularly different from one another, and thus possible candidates for security threats. Since clustering is an unsupervised form of machine learning, the process of splitting data into pieces will be based on the process of training the data cluster and evaluating the results. One may use common cross-validation techniques to ensure that the results are robust. At the same time, hierarchical clustering is another example of unsupervised learning that is aimed at grouping similar objects into clusters. Unlike k-means clustering, it does not require the number of clusters to be defined, but can also be useful for the detection of anomalies and patterns in the security data. As such, the process of splitting data into pieces in hierarchical clustering will also center around the split of the dataset into training and testing sets. The training will focus on the development of a hierarchical cluster structure while the process of testing will evaluate its effectiveness in anomaly detection.

Autoencoder is an unsupervised neural network that is used for this learning and anomaly detection. The main part of it is encoder and decoder, which compress and reconstruct input data, respectively. Anomalies are determined when the reconstruction error is beyond predefined boundaries. The same as with other models, the first stage should be regarding the data splitting. The data should be separated into the training, validation, and testing sets. In general, the set of training data is used to teach an autoencoder model to continue and learn about the shape of the size and shape of the difference from the usual data. The set of samples is used for fine-tuning the model. The set of test examples is used at the end of the implementation to assess how well the model will generalize to new, unseen examples.

4. RESULT AND DISCUSSION

The result is of evaluation of various machine learning algorithms realized during our research offer significant data on their effectiveness to improve the security of a cloud-based educational data storage system. Here is a discussion on the performance of each algorithm and its implications for this study. In this context from Figure 2, logistic regression showed the predictive power with an accuracy of 92%, precision of 88%, recall of 85%, F1 score of 86%, AUC of 0.94, and log loss of 0.25. As such, these results can reveal that logistic regression in our case of application is also efficient for the classification of known threats based on labeled data. Its high accuracy and the great AUC value could indicate that logistic regression is appropriate for distinguishing normal and anomalous behaviors. At the same time, the precision level of 88% that is true for all instances classified as threats is relevant to the concept of the recall value of 85%, which is the proportion of all actual threats. The F1 score of 86% is high, which means that logistic regression is relevant for use in this system.

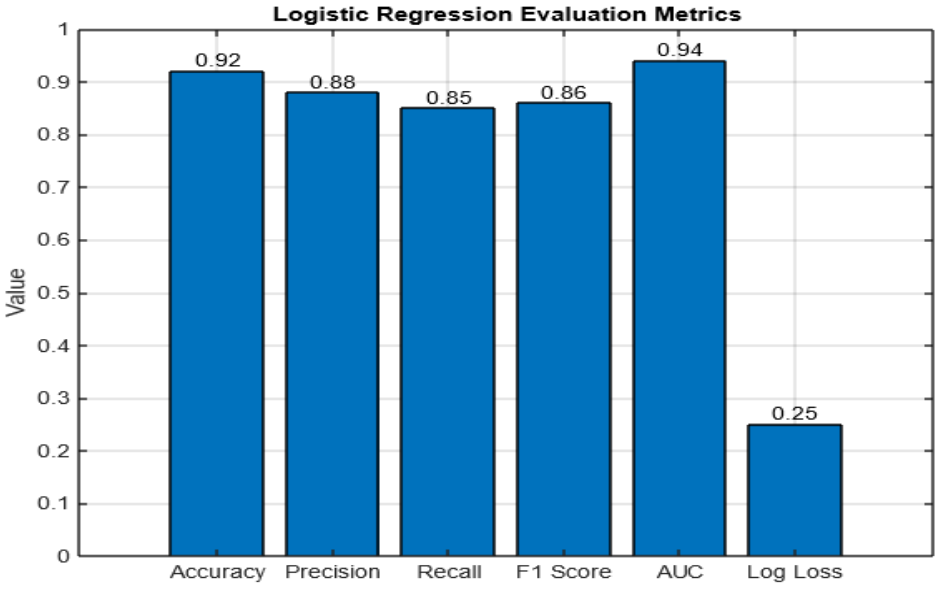

Figure 2: Logistic regression outcome

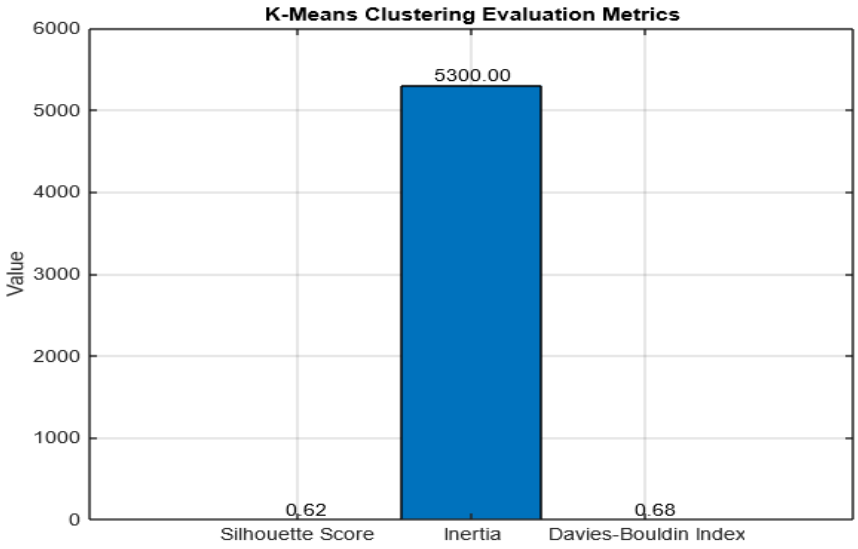

Figure 3: K-means clustering outcome

As it has been mentioned, logistic regression is appropriate for the data where attention should be paid to well-known threats. Nevertheless, this model can be used in the case of an educational setting in which the task is to detect various common security issues, such as attempts to gain access to a firewall or a phishing attack. According to the results, logistic regression can become a satisfactory basic model of detecting threats in the initial stage of such a learning process.

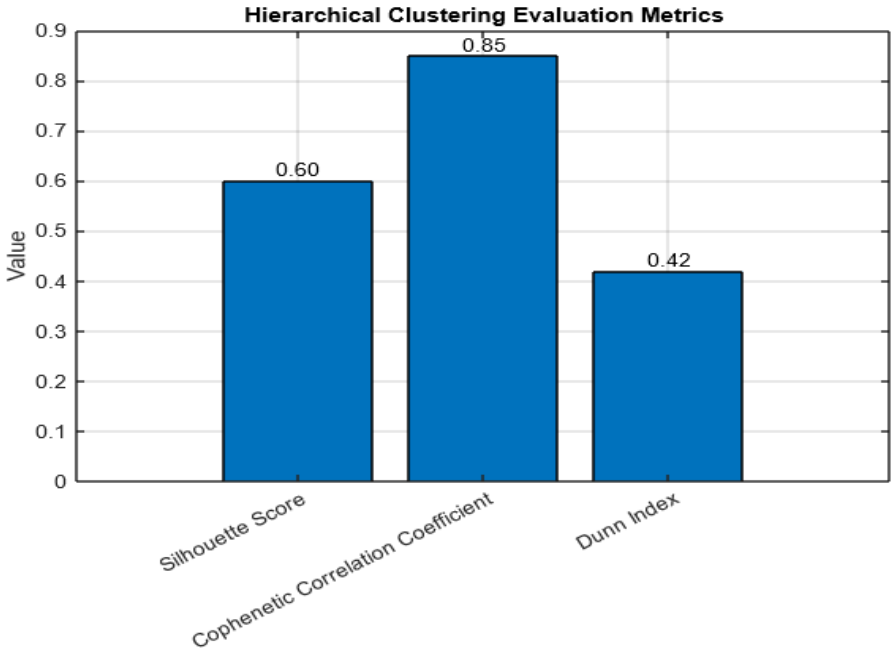

Figure 4: Hierarchical clustering outcome

For k-means clustering presented in Figure 3, the silhouette score was 0.62, inertia was 5300 and Davies-Bouldin index was 0.68. The silhouette score measures how similar an object is to its own cluster compared to other clusters. A score of 0.62 means there is a moderate separation between the clusters. Inertia is the sum of squared distances from each data point to the centroid of its assigned cluster. Therefore, a lower inertia reflects a tighter cluster. Davies-Bouldin index measures the average similarity between each cluster and its nearest neighbouring cluster. Lower values of this index mean better clustering.

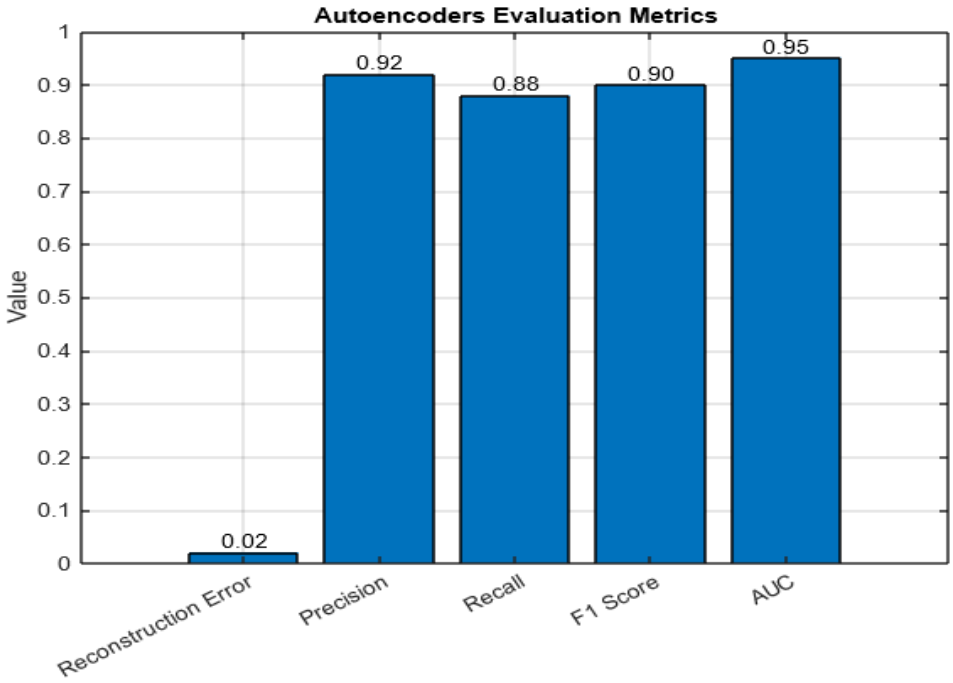

Figure 5: Autoencoders outcome

K-means clustering shows acceptable results, based on the identified groups of anomalous activities. It helps to group all data points into clusters that demonstrate some unusual patterns. In the authentic-replicated experiments, all data points appeared to demonstrate unusual patterns; however, due to the silhouette score of 0.60, some data points might be on the boundary between their clusters and, thus, some unclear results may be obtained in the future. The results of hierarchical clustering were represented in the silhouette score of 0.60, cophenetic correlation coefficient of 0.85, and Dunn index of 0.42. Therefore, the resulting clusters are moderately well-separated as in the case of k-means clustering. The cophenetic correlation coefficient of 0.85 proves good clustering quality. The Dunn index related to clustering compactness and separation was identified at 0.42.

From the hierarchical clustering results in the first section presented in Figure 4, it is possible to gain an insight into the structure and hierarchy of the security threat. In fact, the good cophenetic correlation coefficient and moderate Dunn index indicate that the clusters are meaningful and that this hierarchical structure is well-representative of the relationships between the different types of threats. As such, this method could be used to determine the severity of security threats and understand how different types of security events are interconnected. Regarding autoencoders, it should be noted that the method used to evaluate the performance according to the well-structured hierarchy of reports could not have obtained perfect results. In detail, the reconstruction error of 0.021 suggests that the performance in identifying the differences between the input data and the reconstruction generated by the model in this case was at the least reasonably good. At the same time, based on the precision of 92%, recall of 88% and F1 score of 90%, as well as AUC of 0.95, it is possible to suggest that the models performance in terms of detection of anomalies was higher.

From above Figure 5, Autoencoders perform well when it comes to identifying anomalies in the educational data storage system. The low reconstruction error alongside the high AUC score shows that the model is efficient in detecting deviations in the system. This is vital when it comes to identifying new or previously unseen threats. More still, the precision and recall scores show that the model is good at telling the difference between normal activities and a security breach. From the results of this study, it is evident that a mix of machine learning algorithms, each with their strengths and weaknesses can greatly improve the security of a cloud based educational data storage system. Whereas logistic regression is perfect for known threats, K – means and hierarchical clustering provide valuable insights into the anomalies and patterns present in the data. Autoencoders are always on top due to their ability to identify minute deviations that signal new security threats.

CONCLUSION

Our research has found that machine learning algorithms can be used to protect cloud-based educational data storage systems. According to the findings, logistic regression analysis has produced the following characteristics: 92% accuracy, 88% precision, 85% recall, 86% F1 score, 0.94 AUC, and 0.25-log loss. Based on this data, logistic regression performs highly when classifying known threats. K-means clustering analysis has acquired a 0.62 silhouette score, 5300 inertia, and 0.68 Davies-Bouldin index, which allows it to detect the anomalous clusters. Hierarchical clustering analysis has produced the following indicators: 0.60 silhouette score, 0.85 cophenetic correlation coefficient, and 0.42–Dunn index. Further, the autoencoders' analysis

shows that this method creates a 0.021 reconstruction error and 92% precision, 88% recall, 90% F1 score, and 0.95 AUC.

The results obtained clearly suggest that a range of machine learning approaches are critical in order to detect and resolve security issues at any educational establishment. By taking advantage of such techniques, establishments will be able to continue and monitor data and, in doing so, will be able to ensure that their educational platform remains both safe and intact in the face of emerging cyber threats.

References

1) Radu, I., & Schneider, B. (2023). Computers & Education : X Reality Designing augmented reality for makerspaces : Guidelines , lessons and mitigation strategies from 5 þ years of AR educational projects. *Computers & Education: X Reality*, *2*(May), 100026. https://doi.org/10.1016/j.cexr.2023.100026

2) Sannikov, S., Zhdanov, F., Chebotarev, P., & Rabinovich, P. (2015). Interactive Educational Content Based on Augmented Reality and 3D Visualization. In *Procedia - Procedia Computer Science* (Vol. 66). Elsevier Masson SAS. https://doi.org/10.1016/j.procs.2015.11.082

3) Lawal, K., & Rafsanjani, H. N. (2022). Trends, benefits, risks, and challenges of IoT implementation in residential and commercial buildings. *Energy and Built Environment*, *3*(3), 251–266. https://doi.org/10.1016/j.enbenv.2021.01.009

4) Ho, K., & Tang, D. (2022). A model of behavioral climate change education for higher educational institutions. *Environmental Advances*, *9*(June), 100305. https://doi.org/10.1016/j.envadv.2022.100305

5) Rožanec, J. M., Trajkova, E., Dam, P., Fortuna, B., & Mladenic, D. (2022). Streaming Machine Learning and Online Active Learning for Automated Visual Inspection. *IFAC-PapersOnLine*, *55*(2), 277–282. https://doi.org/10.1016/j.ifacol.2022.04.206

6) He, K., Wang, Z., Li, D., Zhu, F., & Fan, L. (2020). Ultra-reliable MU-MIMO detector based on deep learning for 5G/B5G-enabled IoT. *Physical Communication*, *43*. https://doi.org/10.1016/j.phycom.2020.101181

7) Wood, D. A. (2022). Local integrated air quality predictions from meteorology (2015 to 2020) with machine and deep learning assisted by data mining. *Sustainability Analytics and Modeling*, *2*(October 2021), 100002. https://doi.org/10.1016/j.samod.2021.100002

8) Bermúdez, K., & Caro, K. (2023). *Computers & Education : X Reality Effect of an augmented reality app on academic achievement , motivation , and technology acceptance of university students of a chemistry course*. *2*(August 2022), 1–9. https://doi.org/10.1016/j.cexr.2023.100022

9) Tuli, N., & Mantri, A. (2020). ScienceDirect ScienceDirect Experience Fleming ' s rule in Electromagnetism Using Augmented Reality : Analyzing Impact on Students Learning. *Procedia Computer Science*, *172*(2019), 660–668. https://doi.org/10.1016/j.procs.2020.05.086

10) Clark, R. M., Besterfield-Sacre, M., Shuman, L. J., & Yildirim, T. P. (2008). Work in progress - Assessment of MEA problem solving processes used by engineering students. *Proceedings - Frontiers in Education Conference, FIE*, 37–38. https://doi.org/10.1109/FIE.2008.4720497

11) Farhan, M., Jabbar, S., Aslam, M., Hammoudeh, M., Ahmad, M., Khalid, S., Khan, M., & Han, K. (2018). IoT-based students interaction framework using attention-scoring assessment in eLearning. *Future Generation Computer Systems*, *79*, 909–919. https://doi.org/10.1016/j.future.2017.09.037

12) Bharathi, R., Abirami, T., Dhanasekaran, S., Gupta, D., Khanna, A., Elhoseny, M., & Shankar, K. (2020). Energy efficient clustering with disease diagnosis model for IoT based sustainable healthcare systems. *Sustainable Computing: Informatics and Systems*, *28*(September), 100453. https://doi.org/10.1016/j.suscom.2020.100453

13) Huang, X., Yang, F., Zheng, J., Feng, C., & Zhang, L. (2023). Personalized human resource management via HR analytics and artificial intelligence: Theory and implications. *Asia Pacific Management Review*, *xxxx*. https://doi.org/10.1016/j.apmrv.2023.04.004

14) Ijemaru, G. K., Ang, L. M., & Seng, K. P. (2022). Transformation from IoT to IoV for waste management in smart cities. *Journal of Network and Computer Applications*, *204*(June 2021), 103393. https://doi.org/10.1016/j.jnca.2022.103393

15) Zou, X., O'Hern, S., Ens, B., Coxon, S., Mater, P., Chow, R., Neylan, M., & Vu, H. L. (2021). On-road virtual reality autonomous vehicle (VRAV) simulator: An empirical study on user experience. *Transportation Research Part C: Emerging Technologies*, *126*(February), 103090. https://doi.org/10.1016/j.trc.2021.103090

16) Shankhwar, K., & Smith, S. (2023). Finite element analysis results visualization of manual metal arc welding using an interactive mixed reality-based user interface. *Journal of Manufacturing Processes*, *93*(December 2022), 153–161. https://doi.org/10.1016/j.jmapro.2023.03.001

17) Sutherland, C. A., Albert, W. J., Wrigley, A. T., & Callaghan, J. P. (2008). A validation of a posture matching approach for the determination of 3D cumulative back loads. *Applied Ergonomics*, *39*(2), 199–208. https://doi.org/10.1016/j.apergo.2007.05.004

18) Bouchouia, M. L., Labiod, H., Jelassi, O., Monteuuis, J.-P., Jaballah, W. Ben, Petit, J., & Zhang, Z. (2023). A survey on misbehavior detection for connected and autonomous vehicles. *Vehicular Communications*, *41*, 100586. https://doi.org/10.1016/j.vehcom.2023.100586

19) Cotroneo, D., Simone, L. De, Liguori, P., & Natella, R. (2023). The Journal of Systems & Software Run-time failure detection via non-intrusive event analysis in a large-scale cloud computing platform ☆. *The Journal of Systems & Software*, *198*, 111611. https://doi.org/10.1016/j.jss.2023.111611

20) Segura, E., Morales, R., & Somolinos, J. A. (2018). A strategic analysis of tidal current energy conversion systems in the European Union. *Applied Energy*, *212*(October 2017), 527–551. https://doi.org/10.1016/j.apenergy.2017.12.045

21) Shaffril, H. A. M., Krauss, S. E., & Samsuddin, S. F. (2018). A systematic review on Asian's farmers' adaptation practices towards climate change. *Science of the Total Environment*, *644*, 683–695. https://doi.org/10.1016/j.scitotenv.2018.06.349

22) Nirmal, U., Hashim, J., & Megat Ahmad, M. M. H. (2015). A review on tribological performance of natural fibre polymeric composites. *Tribology International*, *83*, 77–104. https://doi.org/10.1016/j.triboint.2014.11.003

23) Way, M. L., Berndt, N., & Jawad, B. (2003). The study of a cockpit with a fixed steering wheel position: Methods and model. *SAE Technical Papers*, *724*. https://doi.org/10.4271/2003-01-2180

24) Sworna, N. S., Islam, A. K. M. M., Shatabda, S., & Islam, S. (2021). Towards development of IoT-ML driven healthcare systems: A survey. *Journal of Network and Computer Applications*, *196*(June), 103244. https://doi.org/10.1016/j.jnca.2021.103244

25) Pradeep, P., Krishnamoorthy, S., & Vasilakos, A. V. (2021). A holistic approach to a context-aware IoT ecosystem with Adaptive Ubiquitous Middleware. *Pervasive and Mobile Computing*, *72*, 101342. https://doi.org/10.1016/j.pmcj.2021.101342

26) Peixoto, M. L. M., Mota, E., Maia, A. H. O., Lobato, W., Salahuddin, M. A., Boutaba, R., & Villas, L. A. (2023). Ad Hoc Networks FogJam : A Fog Service for Detecting Traffic Congestion in a Continuous Data Stream VANET. *Ad Hoc Networks*, *140*(August 2022), 103046. https://doi.org/10.1016/j.adhoc.2022.103046

27) Hetier, M., Wang, X., Robache, F., Autuori, B., & Morvan, H. (2005). Experimental investigation and modeling of driver's frontal pre-crash postural anticipation. *SAE Technical Papers*, *114*. https://doi.org/10.4271/2005-01-2684