2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC) | 978-1-6654-5401-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/PDGC56933.2022.10053301

Machine learning based data classification methods in cloud security using cloudlightning framework

Tridiv Swain
Kalinga Institute Of Industrial Technology
Bhubaneswar,India
1906446@kiit.ac.in

Awantika Singh
Kalinga Institute Of Industrial Technology
Bhubaneswar,India
1906016@kiit.ac.in

Khushali Verma
Kalinga Institute Of Industrial Technology
Bhubaneswar,India
1906181@kiit.ac.in

Abhay Kumar Sahoo
Kalinga Institute Of Industrial Technology
Bhubaneswar,India
abhaya.sahoofcs@kiit.ac.in

Shefalika Ghosh Samaddar
Dr. Sudhir Chandra Sur Institute of Technology & Sports Complex
India
shefalika.ghoshsamaddar@dsec.ac.in

Rabindra Kumar Barik
Kalinga Institute Of Industrial Technology
Bhubaneswar,India
rabindrafca@kiit.ac.in

Abstract- With the increase in the usage of Cloud services and applications, securing the data naturally becomes a major concern. There are numerous methods in which cloud data can be secured and numerous more ways that are being currently researched. In this paper, we present the approach of data classification and encryption to secure the data. It is incorrect to use security methods without understanding the requirements of that data and hence we use the K-NN classification method to divide the data with respect to their needs followed by encrypting the sensitive data to prevent any security hazards. We used the CloudLightning simulator to test the framework. The framework is broad in scope, but it also aims to enable cloud services for high-performance computing. Infrastructure-as-a-service. The core use case is service provision. However, we believe that genomics, oil and gas exploration, and ray tracing are three downstream use cases that will benefit from the proposed architecture.

Keywords: Cloud Security; CloudLightning; KNN; Cyberinfrastructure as a Service (CaaS); Infrastructure as a Service (IaaS); Platform as a Service (PaaS); Artificial Neural Network (ANN)

I. INTRODUCTION

Nowadays, cloud computing (CC) has become a novel approach for interacting with and delivering services through the Internet.

The normal financial constraints and rising processing costs necessitate the limitation, evaluation, and display of data that have restrained crucial improvements for the current cloud architecture. CC refers to the on-demand opening up of end-client resources, particularly data capacity and handling power, without an instant unique association by the client. The phrase "circulated registering" is well-known and can mean different things to different people. On a single platform across the Internet, appropriate figuring provides the client with both public and private data. In any case, CC has a few security challenges that defer the fast reception of the registering model, like a weakness for clients and affiliations [1, 5].

Application architects and specialized organizations can benefit from distributed processing power at the edge of a framework thanks to edge figuring, an adaption of CC used to handle time-sensitive information. Current edge taking care of grows this methodology through virtualization headway to enhance sending and working to a more wide degree of direction apprehensive servers. The dispersed thought of this perspective presents an adjustment of safety plans used in circled enrollment [2,4].

Furthermore, it is important to accept distinct encryption structures because information may move via a number of Web-related focus points before ending up in the cloud. Additionally, edge focuses may be asset-specific devices, influencing the choice of security protocols. It is possible to transfer the commitment regarding information from professional associations to end clients by maintaining awareness of information at the edges. The core concept is to give PCs the freedom to change on their own, without assistance from humans, and to modify workouts as necessary. In addition to organization types like public, private, local area, and crossover cloud, CC features management models like Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [2-4].

Authorized licensed use limited to: Hochschule Rosenheim (FH). Downloaded on October 21,2023 at 14:51:21 UTC from IEEE Xplore. Restrictions apply.

Reliability, accessibility, and mystery threats are used to describe the significant security stresses in CC. With demands for limitless availability, cloud organizations range from information ability to manage programming organizations. CC is frequently set up as an impenetrable encompassing that may, upon request, provide plan, commitments, and managing force and strength. The cloud model appreciates and supports significant hardware assets for establishment, maintenance, and other purposes. Despite its advantages, CC has problems as another method of calculation. Not every cloud association model is appropriate for every user, provider customer, or complex party. This research uses Machine Learning (ML) computations to illustrate the security concerns and challenges associated with CC and related game plans [3, 4, 5].

ML estimations solve security challenges and direct data. ML employs frameworks to naturally learn and grow without being actively updated. ML is based on the development of computer programs that can locate a suitable velocity and utilize it to determine things on their own. The method of learning begins with experiences or data, such as models, direct understanding, or heading, to redirect for informational structures and choose better choices later in relation to the issue of the presented models. The purpose of this work is to examine the legitimate concerns and security vulnerabilities associated with message handling when ML computations are used. A key factor in turning it into a viable project idea for cutting the cost of both plan and exercises is the creation of affirmation of circled enlisting commitments [2,5]. Similar to this, managing security and assurance risks is crucial in a fluid environment due to the associated concerns with gadget inadequacy. Examining and discussing security concerns that are inextricably linked to ML based circled enlisting with substantial steps taken to legitimate these concerns. A big request for dispersed organizations that are strong, regarded, necessary, and secure was made. Customers who choose their cloud provider carefully currently frequently struggle to find the money to pay any payment for security risk. Additionally, a significant responsibility for cloud framework management providers is security. It is essential to evaluate the security issues faced by cloud supplier sellers' guides and the associated criminal repercussions using various computations. The main topic we examine is security risks in message processing [6].

We currently choose to store all types of data in cloud servers due to limited storage capacity and the need for convenient access. This is also a smart choice for enterprises and organizations to avoid the overhead of deploying and maintaining equipment when data is stored locally. The cloud server provides individuals and organizations with an open and convenient storage platform, but it also brings security issues. A cloud system, for example, may be attacked by both bad users and cloud providers. On these cases, it is critical to secure the security of the data stored in the cloud [10].

We interpret the calculations needed to address problems and forward execution. In addition, we discuss the use of various ML calculations to address security risks in CC. In addition, we suggest research directions that should be considered going forward.

The remainder of the paper is divided into the following sections. Section 2 details the related research work. The cloud lighting framework is covered in Section 3 while the experimental results are analyzed in Section 4. Finally, Section 5 derives the concluding remarks of the present work.

II. RELATED WORK

Encryption must be implemented on a database before outsourcing it to the cloud to prevent any security concerns. Most existing kNN algorithms, on the other hand, are insecure because during query processing they disclose data access patterns. While hiding data access patterns, our technique protects both data and query privacy. In the context of query processing cost, its approach surpassed the existing techniques, according to its performance analysis [1, 4].

As uncertain data is probabilistic in nature [2], searching over it, particularly encrypted uncertain data, remains difficult in theory and practice. To encrypt the object set, we employed modified homomorphic encryption, which enables for addition and multiplication over encrypted data. The predicted rank can be utilized as a ranking criterion with confidence because it satisfies the critical top-k properties. To demonstrate the efficacy of the suggested approach, the researchers used both real-world and synthetic datasets to evaluate its performance. With the help of Training dataset Filtration Key Nearest Neighbor (TsF-KNN) classifier, the goal of this research [3] is to demonstrate getting portable information stockpiling mystery and protection in cloud correspondence structure as far as programmed information characterization using flexible preparing datasets.

It groups the information based on the record's classification level with higher accuracy and a strong timetable than traditional K-NN computations, and getting such private information classification later by applying existing cryptographic solutions for ensuring information security and secrecy levels. Reenactment results show that reducing the overall cost and limiting procedural time, increasing system performance, and securing the system are possible. In [4], a configurable, high-performance KNN accelerator called CHIPKNN is designed and implemented. This program may automatically design the optimal KNN accelerator architecture that exceeds the off-chip memory bandwidth limit or the FPGA resource limit given key KNN parameters and a target cloud FPGA platform. Tuning the memory access harborage range, successive data access size, and concurrent number of memory banks improves off-chip memory access. Our robotization tool uses a logical

Authorized licensed use limited to: Hochschule Rosenheim (FH). Downloaded on October 21,2023 at 14:51:21 UTC from IEEE Xplore. Restrictions apply.

performance model to discover the ideal design with balanced execution of all steps.

It is obvious that cloud computing is vulnerable to a variety of attacks and threats from both internal and external sources [11]. However, AI-based technologically advanced technologies that ensure cloud security and secure data transfer across the network are advantageous in this regard. Artificial Neural Network (ANN) techniques can be used to assess the performance of cloud security networks. This tool assures data security while also protecting workload and facilitating data transport. ANN aids in the detection of attacks and threats at various levels of the cloud network, increasing the overall system's reliability.

Cloud computing [5,7], which is the most recent development, can accommodate and serve multiple customers and services on-demand. Machine learning techniques are used to discover failures in this study, and proactive fault tolerance strategies are used. Different workflow topologies were used to test the proposed system. The k-means pipeline environment has the highest success rate compared to others. Future discussions will focus on federated cloud reactive fault tolerance. This survey analyzed security issues and assaults as CC's biggest problems. ANNs, K-NN, Nave Bayes, SVM, K-Means, and SVD were studied as CC security solutions [12-15]. CC models are based on future assessment headings. CC is a collection of public and private data centers that offer a single Internet platform. We assess administered, solo, semi-managed, and learning ML calculations used to beat cloud security concerns. Then, we consider each method's elements, merits, and drawbacks.

III. PROPOSED FRAMEWORK

In order to work on the reliability of logical work processes, we have implemented a potent bunching method and the concept of association titles. By considering the already available cloud assets, it increases the failure rate. In this study, we suggest improving internal failure adaptation for logical work processes using calculations that try to redo the task using spatial worldly data. Propose using solo learning models to implement adaptation to internal failure for extensive logical work process applications.

We suggest a simple check-guiding method for identifying failed projects and estimating the number of replications using heuristics. The majority of the results for further developing adaptability to non-critical failure in logical work process applications in cloud stages focus on performing checkpointing or computations for replication of tasks to reduce cost or turnaround time.

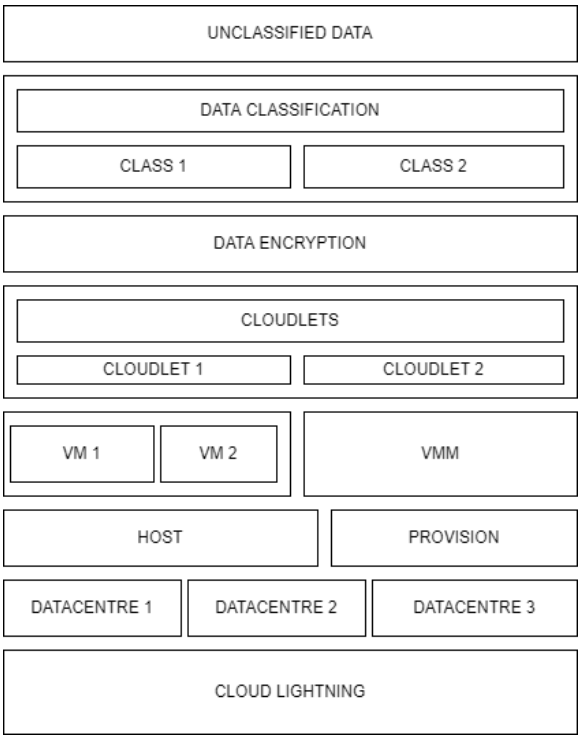

Fig. 1. Represents the cloudligtning framework

Psedocode For CloudLightning Framework:

Input : Layer and Cloudlightning_number

1: create model node
2: if Layer[Cloudlightning_number].ID =1 then
3:production.rule = High clouds rule
4:if Layer[Cloudlightning_number].ID = 2 then
5:production.rule = Middle clouds rule
6:if Layer[Cloudlightning_number].ID = 3 then
7:productionrule = Low clouds rule
8:endif
9:L-system cloudLightning modelling using production_rule
10: return model

Figure 1 depicts the framework, which describes a method for addressing security issues in the cloud. For simulation, the CloudLightning simulator was employed. Since it offers the most fundamental organizational structures, the unclassified data was initially sent for data classification. In order to identify potential dangers, it also helps to group the data according to their sensitivity.

After that, the data is encrypted and split up into cloudlets. As data is transmitted to and from the cloud

Authorized licensed use limited to: Hochschule Rosenheim (FH). Downloaded on October 21,2023 at 14:51:21 UTC from IEEE Xplore. Restrictions apply.

application or network, encryption will help to keep it safe. Cloudlets were assigned to and managed by VMs using the Virtual Machine Manager (VMM). Then they are hosted, making websites or applications accessible and higher scalability and flexibility are provided because they are housed on actual cloud servers or a network that is virtually connected. At the very end, simulation is performed using CloudLightning because it offers simulation of both conventional and self-organizing clouds.

The proposed work has as its goal performing various AI calculations for failure expectancies for various work process applications. Additionally, failure is common in the cloud stage, and in this work job disappointment forecast is finished by the expectation unit in the suggested framework. Various AI computations are also executed but-centrically. AI techniques like Decision Tree, K-Means, and KNN Algorithm are applied to the model obtained from the operational work processes of various constructions in order to accurately predict assignment failures. Different assessment measures are used to examine and check the suggested models' accuracy. Table 1 compares the different frameworks in terms of their research, limitations, advantageous and technologies used.

Table 1. Table represents comparative study of framework [7-9]

RESEARCH	**LIMITATION**	**ADVANTAGES**	**TECHNOLOGY**
Beyond lightning: A survey on security challenges in cloud computing	security in cloud computing	reduce operating costs while increasing efficiency	Cloud SLA
Towards enabling Cyberinfrastructure as a Service in Clouds	Security in CaaS	a lightweight middleware for access and manage modern cyber infrastructures	Cyberinfrastructure as a Service (CaaS)
Cloud Lightning	Limited	Provides live implementation in a more secure way.	CloudLightning
Container-Based Virtualization for Heterogeneous HPC Clouds: Insights from the CloudLightning Project	Resource utilization	performance portability, scalability, as well as enable a variety of HPC	HPC,Cloud Lightning

IV. EXPERIMENTAL RESULT ANALYSIS

Cloud computing (CC) is a relatively new framework for collaborating with and communicating with administrations through the Internet. The normal financial constraints and rising computing costs necessitate data limitation, assessment, and display, which have restrained vital adjustments for the current cloud architecture. CC refers to the openness of end-client assets, notably data capacity and processing power, on demand and without the client's immediate unique association. Circulated registering is a well-known expression that means different things to different people. Appropriate calculating provides clients with public and private data on a same platform over the Internet. In any event, CC has a few security issues that are delaying the adoption of the registration paradigm, such as vulnerability for clients and affiliations. Table 2 describes about the study of algorithms and their accuracy.

Table 2. Study of Algorithm versus Accuracy

ALGORITHM	**ACCURACY**
KNN	94.5%
CHIP-KNN	91%
SVM	78%
K-MEANS	83%

Authorized licensed use limited to: Hochschule Rosenheim (FH). Downloaded on October 21,2023 at 14:51:21 UTC from IEEE Xplore. Restrictions apply.

The goal of this analysis is to compare the performance of each algorithm rather than to determine the best performance for each. According to the result analysis graph, the KNN algorithm can be selected as the most suitable one due to its high accuracy, which is 94.5%, because KNN can analyze even very complex patterns and hence more accurate predictions. This is not the case for the SVM algorithm which shows the lowest accuracy of all. Figure 2 visualizes the different result analysis of algorithm used with respect to their accuracy.

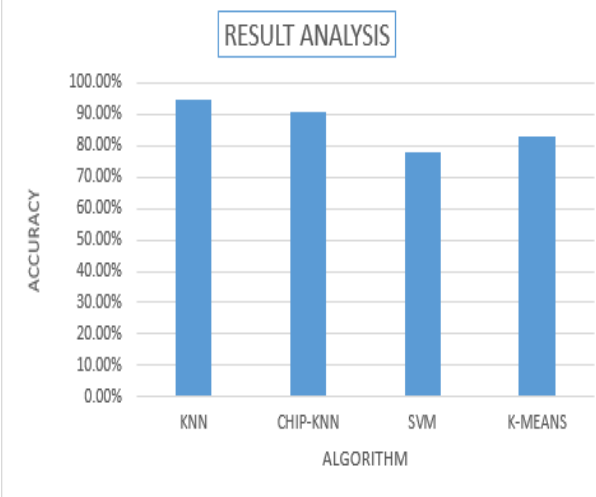

Fig. 2. Result Analysis of algorithm used

Table 3. Risk analysis of algorithm used

ALGORITHM	**RISK PERCENTAGE**
K-MEANS	87%
KNN	45%
CHIP-KNN	23%
SVM	31%

Table 3 represents the risk analysis of algorithms and their risk percentages. Algorithmic risks emerge from the utilization of information investigation and cognitive innovation based programming calculations in different computerized and semi-mechanized dynamic conditions.

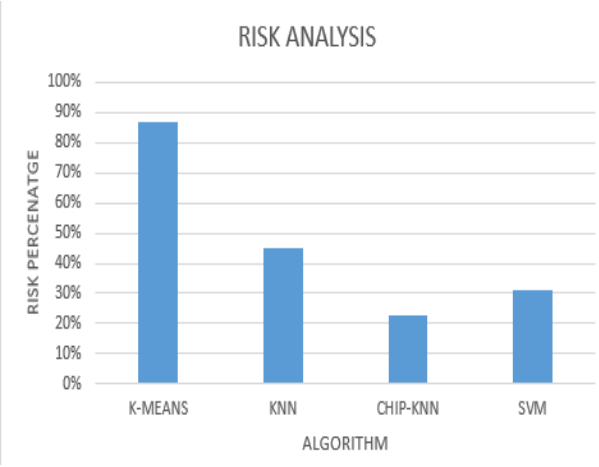

Fig. 3. Risk factor of algorithms used

.

Figure 3 gives a framework to understanding the various regions that are weak against such risks and the basic elements causing them. From the overview of the result we can easily identify the most significant and non-significant risk factor. In CHIP-KNN and SVM, the model can be analyzed with the most significant risk factor. So for cloudlightning, KNN gives the best accuracy of 94.5%.

V. CONCLUSIONS

In this research, we proposed a framework to increase the security of the data in the cloud. In this research, we describe a method of data classification and encryption for data security. It is inappropriate to utilize security measures without first understanding the demands of the data, thus we apply the K-NN classification method to partition the data according to their needs before encrypting the sensitive data to avoid any security risks. To test the framework, we used the CloudLightning emulator. More studies and research should be done to prevent mislabeling of data during classification which is one of the most commonly made mistakes.

REFERENCES

[1] Kim, H. J., Lee, H. J., & Chang, J. W. (2021, January). A secure and efficient query processing algorithm over encrypted database in cloud computing. In 2021 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 219-225). IEEE.

[2] Guo, C., Zhuang, R., Su, C., Liu, C. Z., & Choo, K. K. R. (2019). Secure and efficient K nearest neighbor query over encrypted uncertain data in cloud-IoT ecosystem. IEEE Internet of Things Journal, 6(6), 9868-9879.

[3] Inani, A., Verma, C., & Jain, S. (2019, February). A machine learning algorithm TSF k-Nn based on automated data classification for securing mobile cloud computing model. In 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS) (pp. 9-13). IEEE.

Authorized licensed use limited to: Hochschule Rosenheim (FH). Downloaded on October 21,2023 at 14:51:21 UTC from IEEE Xplore. Restrictions apply.

[4] Lu, A., Fang, Z., Farahpour, N., & Shannon, L. (2020, December). CHIP-KNN: A configurable and high-performance k-nearest neighbors accelerator on cloud FPGAs. In 2020 International Conference on Field-Programmable Technology (ICFPT) (pp. 139-147). IEEE.

[5] Prathibha, S. (2019, December). Investigating the Performance of Machine Learning Algorithms for Improving Fault Tolerance for Large Scale Workflow Applications in Cloud Computing. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 187-190). IEEE.

[6] Liu, Q., Hao, Z., Peng, Y., Jiang, H., Wu, J., Peng, T & Zhang, S. (2021). SecVKQ: Secure and verifiable kNN queries in sensor–cloud systems. Journal of Systems Architecture, 120, 102300.

[7] Lynn, T., Xiong, H., Dong, D., Momani, B., Gravvanis, G. A., Filelis-Papadopoulos, C & Morrison, J. P. (2016). CLOUDLIGHTNING: A Framework for a Self-organising and Self-managing Heterogeneous Cloud.

[8] Khan, M., Becker, T., Kuppuudaiyar, P, & Elster, A. C. (2018, April). Container-based virtualization for heterogeneous HPC clouds: insights from the EU H2020 cloudlightning project. In 2018 IEEE International Conference on Cloud Engineering (IC2E) (pp. 392-397). IEEE.

[9] Zardari, M. A., Jung, & Zakaria, N. (2014, June). K-NN classifier for data confidentiality in cloud computing. In 2014 International Conference on Computer and Information Sciences (ICCOINS) (pp. 1-6). ieee.

[10] Shen, J., Zhou, He, D., Zhang, Y, Sun, X, & Xiang, Y. (2017). Block design-based key agreement for group data sharing in cloud computing. IEEE Transactions on Dependable and Secure Computing, 16(6),996-1010.

[11] Tadeo, D. A. G., John, S. F., Bhaumik, A., Neware, R., Yamsani, N., & Kapila, D. (2021, December). Empirical Analysis of Security Enabled Cloud Computing Strategy Using Artificial Intelligence. In 2021 International Conference on Computing Sciences (ICCS) (pp. 83-85). IEEE.

[12] Sahoo, A. K., Pradhan, C., & Das, H. (2020). Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making. In Nature inspired computing for data science (pp. 201-212). Springer, Cham.

[13] Sahoo, A. K., Raj, S., Pradhan, C., Mishra, B. S. P., Barik, R. K., & Vidyarthi, A. (2022). Perturbation-Based Fuzzified K-Mode Clustering Method for Privacy Preserving Recommender System. International Journal of Information Security and Privacy (IJISP), 16(1), 1-20.

[14] Mishra, B. K., Sahoo, A. K., & Pradhan, C. (2018). GPU based reduce approach for computing faculty performance evaluation process using classification technique in opinion mining. International Journal of Data Analysis Techniques and Strategies, 10(3), 208-222.

[15] Sahoo, A. K., Pradhan, C., & Mishra, B. S. P. (2019, April). SVD based privacy preserving recommendation model using optimized hybrid item-based collaborative filtering. In 2019 international conference on communication and signal processing (ICCSP) (pp. 0294-0298). IEEE.

Authorized licensed use limited to: Hochschule Rosenheim (FH). Downloaded on October 21,2023 at 14:51:21 UTC from IEEE Xplore. Restrictions apply.