Establishment of Systematic Data Security in Cloud using Machine Learning

RAKSHITHA S[1], SHWETHA N[2], THANUSHA S P[3], SHRIHARI M R[4]

[1-3]UG Student, Dept. of Computer science Engineering, SJC Institute of technology, Chickballapur, Karnataka, India
[4]Assistant Professor, Dept. of Computer science Engineering, SJC Institute of technology, Chickballapur, Karnataka, India

Abstract – *A data contribution in the cloud is a strategy to permit clients to promisingly give a correction of passage to data or an information in excess of the cloud. Data holder can't oversee over their data, since cloud assessment donor is an outsider benefactor. The main tragedy with data that take part in the cloud is the isolation and secure measure issues. Different approaches are acquiring to support user seclusion and protected data. Security should be provided to store any kind of data and storing that data in a cost-effective manner is also important so cloud has been used. The data holder reevaluates their information in the cloud because of cost diminishing and the tremendous assets given by cloud administrations. This paper point of convergence on various plans to decrease through ensured information sharing like data commitment with forward security, protected information partaking for vigorous gatherings, quality-based data sharing, scrambled information sharing and normal effect Based Privacy-Preserving check set of rules for right to use regulate of rethought information. Machine learning classification algorithms such as decision tree, Support Vector Machine (SVM), and Logical regression are implemented to predict the performance.*

Key Words*: *Security, Logical regression, Third party, Decision Tree, Machine learning techniques, Internet Protocol, Wi-Fi, Mobile Application.

1.INTRODUCTION

Cloud computing is a technological advance that offers the facilities, platform & software of information technology as internet services. The prevalence and use of Cloud registering are expanding quickly. A few organizations are putting resources into this field either for their own utilization or to give it as a help to other people. One of the aftereffects of Cloud advancement is the rise of different security issues for both industry and customer. Providing data security is utmost important for storing any organizations data, security threat is a major concern these days. The principle objective is to allow clients to utilize and pay for what they need, promising on-request benefits for their product or framework needs. It is considered to be the conversion of a long-lasting dream called Computing for One of the aftereffects of Cloud advancement is the rise of different security issues for both industry and customer. Many organizations always have the problem to store and secure the data. Especially when it comes to securing students data like personal data ranking, academic performance of each and every student.

Distributed computing is a seen of huge and positive IT framework shift, to limit its inadequacies much security work is expected. Despite the large number of research studies conducted on Cloud security using machine learning, the ML techniques used for Cloud security, the security areas that ML techniques are used for, and the estimation and accuracy of the ML techniques are utilized.

Machine learning is a type of artificial intelligence that allows software applications to become more perfect at predicting outcomes without being distinctly programmed to do so. ML algorithms use ancient data as input to predict new output values. Machine Learning is used in many ways for cloud attack detection. It detects & attacks users when an attack happens & it prevents the attack before it happens by checking the security itself for any vulnerabilities. Machine learning techniques are very helpful for identifying attacks & it includes a series of algorithm that can learn patterns from data and predict accordingly. At present, machine learning algorithms are most popular to evaluate data that has been extensively applied in the education sector. Security to data has been given using AES encryption algorithm which helps is protecting it from malicious users. Security necessities significantly rise while putting away close to home recognizable on cloud climate. The data is then stored to cloud since it is easier to operate and cost-effective.

2. RELATED WORK

The results from paper [1] Cloud computing (CC) is a developing pattern in numerous fields like IT areas, clinics, finance on account of the compelling use of assets through the arrangements. Each advancing execution in the cloud faces numerous difficulties like protection and security of client's information. In a unified climate, the information can be adjusted without the information on proprietor by unapproved clients (i.e., security break is unavoidable).

[2] This paper centers around the plan and improvement of an API Gateway, which gives an extension between end-clients and their information sources, and the C3ISP

Framework. It smooths the way for end-users to retrieve their CTI data, and regulate data sharing agreements in order to sterilize the data. The outcome of these tests will show the productiveness of our entryway plan, and the advantages for the end-users who will use it to get to the C3ISP substructure.

[3] Providing forward protected ID based ring signature method security level of increased ring signature. If secret key of any user has been compromised, previous generated signatures of all is included and the user still remains valid. If a secret key of a user has been compromised it is impossible to ask all data owners to reauthenticate their data. Providing forward secure ID based ring signature method security level of ring signature increased. If secret key of any user has been compromised, previous generated signatures of all is included and the user still remains valid. If a secret key of a user has been compromised it is impossible to ask all data owners to reauthenticate their data.

[4] An attribute based secure data sharing scheme with EABDS in cloud computing. The data confidentiality & to achieve fine-grained access control this scheme encrypts data with DEK using symmetric encryption method & then encrypts it in perspective of CP-ABE. The homomorphic encryption is used to solve key escrow problem in order to generate attribute secret keys of users by attribute authority in support with key server. EABDS conspire accomplishes quick characteristic renouncement which certifications forward and in reverse security, and less calculation cost on clients. Benefits of this strategy are safer and proficient.

[5] A typical authority-based insurance saving approval show address security issues for a conveyed stockpiling. This attracts for multi-user collaborative cloud applications. Security arrangements essentially center around verification. The SAPA shared access authority is achieved by anonymous access request matching mechanism, provides Ciphertext-policy attribute-based access control to empower clients to dependably get to its own information fields and intermediary re-encryption is applied to give information dividing between various clients. This addresses user's sensitive access related privacy during data sharing in cloud environment and achieves data access control, access authority sharing & privacy continuation shielding. SAPA convention, verification and approval is protected without compromising client's private data.

[6] Predicting the Student Academic Performance in Knowledge, Skills and Abilities (KSA) using Data Mining Techniques and also by using Machine Learning Techniques. The main objective of this paper here is to be implemented in higher education institutions like schools and colleges that is to provide a quality education to its students. One approach to accomplish most elevated level of value is to distinguish factors influencing scholarly execution and afterward attempting to determine shortcoming of these building blocks development. The algorithm used here is classifying and segregating the data successfully in to two clusters by using the features mentioned above that will confirms the appropriateness of the selected attributes for a forecasting cause.

3. METHODOLOGY

The proposed approach is works on as, while data is from the user, the database will be created if it is not present in a directory with tables which are familiar. One table is for the data and IP address of the user which is sent and another table is about user details with credentials of user. The verification of the user is done by IP address and email. Logistic Regression is a mathematical modelling process which describes the relationship between several independent variables, X1... XK, and a reliant variable, D. The strategic model uses the calculated capacity as a numerical structure which has the reach somewhere in the range of 0 and 1 for some random information.

Decision Tree: A decision tree model represents a tree structure that is similar to a flowchart. In this design, each interior hub addresses a test on a dataset property while each tree limb addresses the test result.

Linear Regression Algorithm: Linear Regression Algorithm is an AI calculation dependent on regulated learning. Linear regression algorithm is a part of regression analysis. Regression analysis is a method of prescient demonstrating that assists you with discovering the connection among Input and the objective variable.

AES: Advanced Encryption Standard is also called as Rijndael algorithm. It is symmetrical block cipher algorithm that will converts to cipher text from taken plain text in blocks of 128 bits using keys of 128, 192, and 256 bits anda byte oriented. The structure depends on a replacement change organization. The determination cycle for this is secret yet acknowledged open public remark.

DES: Data Encryption Standard, it is a block cipher code calculation that takes plain content in squares of 64 pieces and converts them to ciphertext utilizing keys of 48 bits. It is a symmetric key algorithm and is a Bit-Oriented. The determination cycle for this is secret. Known assaults against DES incorporate Brute-power, Linear grave investigation, and Differential sepulcher examination.

To meet the needs of clients the cloud computing will provides a service. This makes the cloud critical as individuals begin to rely upon it.

Machine Learning is used in many ways for cloud attack detection. It detects the users when an attack happens & it prevents the attack before it happens by checking the security itself for any vulnerabilities.

This assists with creating significant meta-psychological abilities that add to a scope of significant alumni capacities. All experts should have the option to assess their own exhibition, so this training ought to be implanted in advanced education learning as ahead of schedule as could really be expected.

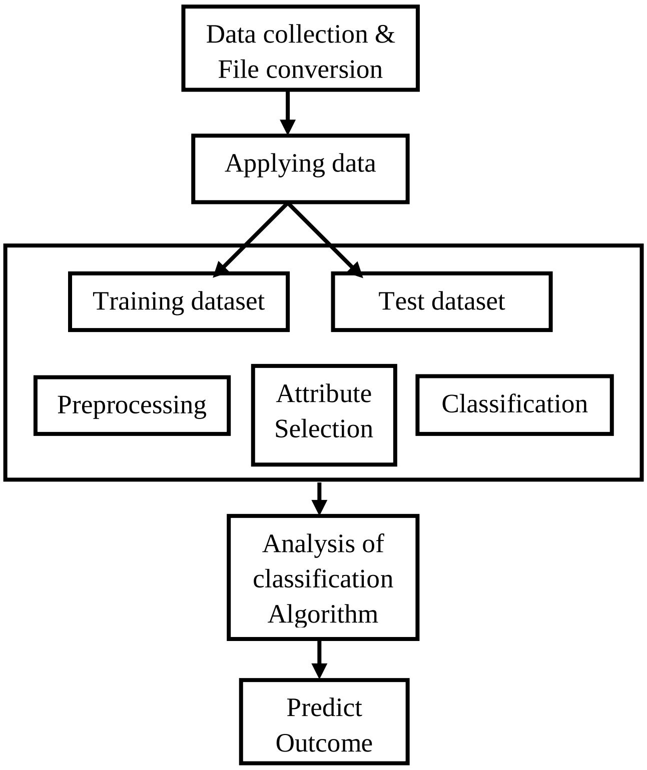


Figure 1. System Framework

The objectives are as follows,

- To study how to store data with high security using encryption methods in Machine Learning.
- The different ML algorithms are used to overcome the cloud security issues, which compares the performance of each technique based on their features.
- To implement advanced concepts of ML to segregate and classify data.
- To execute capable data security.
- AES/DES algorithms are used for securing the data.
- To store information in Cloud which is cost proficient .
- To design the system which is efficient in cost and maintenance.

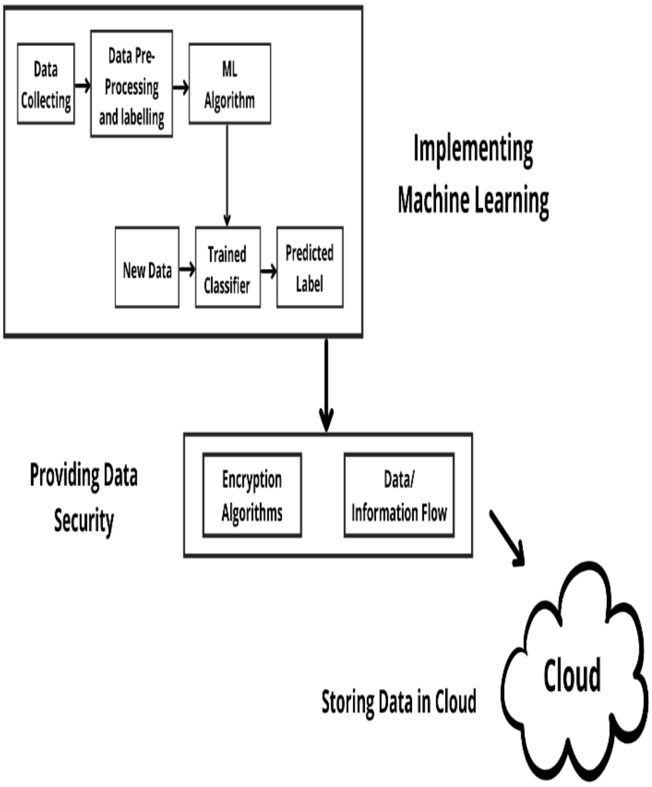


Figure 2. Flow Diagram

Algorithm

1. Collecting the data from the data sources.
2. Preprocessing the data in order to get a normalized dataset and then labelling the data rows.
3. The aftereffect of the subsequent advance, the preparation and testing dataset, is taken care of to the Machine Learning Algorithm.
4. The ML algorithm builds a model using the training data and test the model using the test data.
5. The ML algorithm produces a trained model or trained classifier that can take as an input a new data row and predicts its label.
6. Utilizing AES encryption calculation, the gathered information is ensured and gotten that is protected and secured.
7. The Secured data is stored in a cloud/local server.

Flow chart

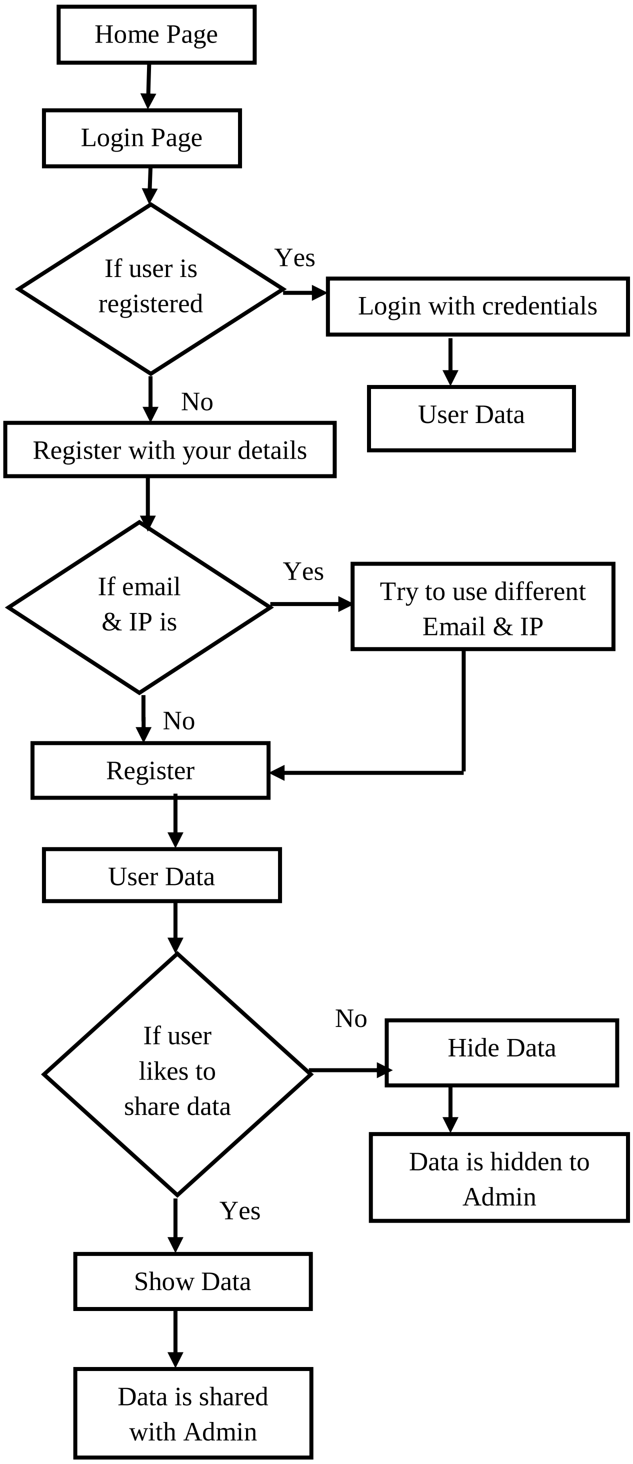


Figure 3. Flow Chart

4. PROPOSED SYSTEM

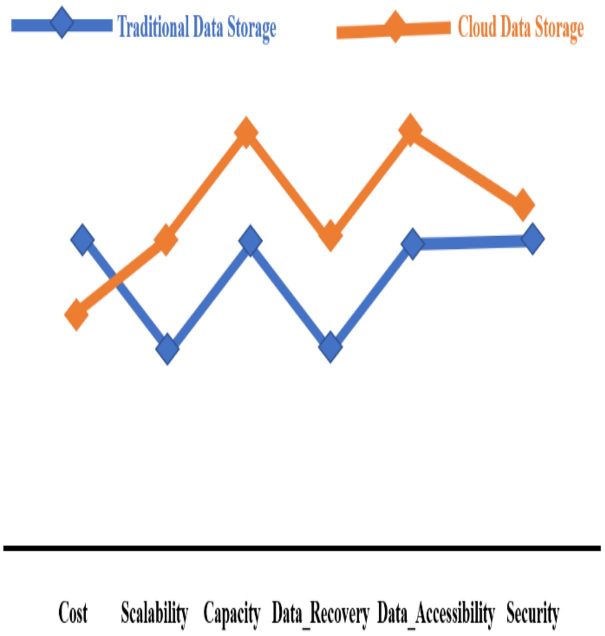
The goal of our project is we build a privacy-aware framework to secure private data and make the user aware of what data is being shared. The data is pulled from file using get API and post API where a token is generated and the data is obtained in json format. We used the TCP/UDP tool to send the data from the phone to the server. The information is sent in word reference design i.e., dictionary format. And now user data will be collected from the clients mobile to server and a prompt will pop up by questioning to store the data in the database or not and also asks the user to set expiry of data. If we click yes, on the off chance that we click indeed, the information will be shipped off the data set in an AES encoded design or on the other hand in case no is chosen by the client it will consequently kill the interaction.

We created an application for the user to login with their credentials and see what data is stored and there is a registration form details for any new user who wants to store an information or data. When the user gets logged in, they can see the data which is stored by only his concern. There is an option in the login page that users can set which data can be viewed by admin and which admin can't viewed only the user has the privilege to show the data and hide the input provided by user.

5. RESULT ANALYSIS

The socket is created and then the port is reserved for a service and then host will bind. After the above process the connection is made with the client and after sending the data from the client to the data is decrypted by fetching the IP data and loads to the set of information. Traditional storage requires physical drives to share data and network is to establish between both. In this framework document access time is subject to the organization speed This framework has quick access time when contrasted with distributed cloud storage. Cloud storage are more secure as it integrates with security tools. So, by these all the data will be stored securely.

Name	Email	Phone no	Gender	Address
John	Joh6@gmail.com	9632587412	Male	abc
Mary	Mar1@gmail.com	7456982156	Female	xyz
Rani	Ra8@gmail.com	8521479632	Female	pqr



Traditional v/s Cloud Data Storage

3. CONCLUSIONS AND FUTURE WORK

The main aim is to securely store and access data which is not controlled by the owner of the data in cloud. We exploit the encryption technique to protect data files in the cloud. The algorithms that are used here for encryption will be improve the performance during encryption and decryption process. This method of putting away and getting to information is much get and have superior. Here the is analyzed and segregated with various machine learning algorithms. The classification algorithms are used frequently. Decision tree, Logical Regression algorithms are used here for these purposes.

The new applications are producing immense measure of information in organized and unstructured structure. To store and process the Big data can be used and process the data and probably more amounts in near coming future. Ideally, Hadoop will improve and it is better. New advancements and devices that have capacity to record, screen measure and join a wide range of information around us, will be presented soon. We will need new technologies and tools for anonymizing data, analysis, tracking and auditing information, sharing and managing, our own personal data in future. Such countless parts of the life are wellbeing, schooling, telecom, showcasing, sports, and business etc., that regulates extremely colossal data world ought to be cleaned in future.

REFERENCES

[1] Brijesh Kumar Baradwaj, and Saurabh Pal, "Mining instructive information or a data to investigate understudies' exhibition". International Journal of Advanced Computer Science and Applications Vol. 2, No. 6. (2018).

[2] Md. Hedayetul Islam Shovon and Mahfuza Haque, “An Approach of Improving Academic Performance of students by utilizing K-means clustering algorithm and Decision tree”, International Journal of Advanced Computer Science, (2015).

[3] Carlos Márquez-Vera, Alberto Cano, Cristóbal Romero & Sebastián Ventura, “Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data”, Applied Intelligence, Vol.38, (2013), pp.315-330.

[4] Thaddeus Matundura Ogwoka, Wilson Cheruiyot, and George Okeyo, “A Model for Predicting the Academic Performance of Students by using a Hybrid of Decision tree Algorithm and K-means Algorithm”, International Journal of Computer Applications Technology and Research, Vol. 4, No. 9, (2020), pp.693 – 697.

[5] Sen, J. (2011c) “A Secure and Efficient way of Searching for Trusted Nodes in Peer-to-Peer Network”. The events on Computational Intelligence in Security for Information Systems (CISIS’11) of the 4th International Conference, pp. 101-109, Springer LNCS Vol 6694, June 2018.

[6] Nawal Ali Yassein, Rasha Gaffer M Helali and Somia B Mohomad , “Predicting Academic Performance of students in KSA using Data Mining Techniques”, Journal of Information Technology & Software Engineering., Vol.7, No. 5, (2020).

[7] Xinyi Huang et.al "Efficient and secure character-based encryption plot with balance test in distributed computing," (2015).

[8] Huang Qinlong et.al “A scalable attributed-set-based access control with both sharing and full-fledged delegation of access privileges in cloud computing," (2015).

[9] Hong Liu et.al “In the multi-authority cloud Securing outsourced data with fine-grained access control and efficient attribute revocation," (2015).

[10] Xin dong et.al “Privacy-preserving public auditing for secure cloud storage,” (2014).

[11] Qiang Tang et.al “Secure Multi-Authority Data Access Control Scheme in Cloud Storage System Based on Attribute-based Sign encryption" (2014).