International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Data Mining Techniques for Cloud Privacy Preservation

Ghaith Mousa Hamzah Amlak [1], & Karim Hashim Kraidi Al-Saedi [2]

Submitted: 12/02/2023 **Revised:** 18/04/2023 **Accepted:** 09/05/2023

Abstract: A high standard of phishing prevention became essential when it came to Internet phishing. As a result of sophisticated phishing attacks, a new mitigation challenge was created. Internet phishing has recently raised serious security and financial issues for people and businesses around. There has been a significant financial loss associated with internet services that provide a variety of communication channels, including electronic commerce, online banking, research, and online trading, as well as those that exploit both software and human weaknesses. The aim of the study is to identify the extent of works and their limitations for cloud security, to design and develop the solution for cloud security, and to evaluate the model.
We propose this new model mining model, which consists of two main bases: preprocessing and classification. the accuracy obtained in this study for malware analysis is quite high, with the highest accuracy being 95.2%.
The results of the Intrusion features, Features of Models, Malware Detection Graphs for the data, Attacks Identified from different countries for the data, Machine learning Based Model Implementation, and Accuracy of analysis were analyzed and data were obtained. Deploying sensors or agents on cloud-based systems and apps is often the first step in the data collection process for malware investigation in the cloud. Network-based intrusion detection systems (NIDS) or intrusion prevention systems are a popular way to collect data. An important paradigm that delivers the results for the grouping of requests is the application of SVM to the problem of assault detection.

Keywords*: security, cloud computing, data mining, phishing.*

1. Introduction

Internet phishing prevention on the highest level became essential. A new mitigation difficulty was generated by the intimidation caused by more sophisticated phishing assaults [1,2]. Internet phishing has recently raised serious security and financial issues for people and businesses around. Internet services that provide a variety of communication channels, including electronic commerce, online banking, research, and online trading, and that take advantage of both software and human weaknesses suffered from significant financial loss [3,4]. As a result, improved privacy-preserving data mining techniques are constantly needed for online information exchange that is trustworthy and secure. Data mining algorithms became significantly more complex as a result of the huge expansion in consumer personal data storage, which had a big impact on information exchange [5,6].

The Privacy Preserving Data Mining (PPDM) algorithm, among other existing algorithms, produces great results in terms of the inner perception of preserving privacy and data mining. The three mining aspects—association rules, classification, and clustering—must all be protected in terms of privacy [7,8]. Many communities, including the database, statistical disclosure control, and cryptography communities, have engaged in an extensive discussion of the issues raised by data mining. The development of new cloud computing technology made it possible for business partners to exchange information and data for mutual advantage. All of these are connected to the growing complexity of data mining techniques that have an impact on information sharing, as well as the cumulative ability to store users' specific data [9-11]

Several privacy protection techniques for data mining are currently available. These include cryptography, L-diverse, taxonomy tree, randomization, clustering, association rule, and K-anonymity [12]. They also include classification, distributed privacy preservation, association rule, clustering, and taxonomy tree. The PPDM methods secure the data by altering it to cover up or remove the original, sensitive information. They often rest on the ideas of privacy failure, the ability to distinguish between original user data and updated data, information leakage, and assessment of data accuracy loss [13,14]. These methods' primary goal is to provide a trade-off between precision and privacy. Other strategies that use cryptographic methods to stop data leakage are very costly computationally. On the other hand, PPDMs make advantage of data distribution and partitioning that

1 kufa university ,ph.D in computer science and mathematics college ,iraq
2 Mustansiriyah University, College of Science, Department of Computer
Tammar2009ghaith@gmail.com
dr.karim@uomustansiriyah.edu.iq

is dispersed horizontally or vertically among numerous organizations [19-22].

Sometimes people are hesitant to disclose the entire set of data and may want to use different protocols to block the information [23]. The main justification for using such methods is to protect people's privacy while obtaining overall results from all the data. Despite extensive research, a technique with acceptable privacy settings is still a long way off. Before data is transmitted to several cloud service providers, it must be protected [24]. Prior to exchanging customer data with unauthorized third parties who are not directly permitted access, it is necessary to identify the clients in order to protect their privacy. This can be accomplished by removing the unique identifying fields from the dataset, such as name and passport number [25]. Despite the elimination of some information, other pieces of information, such as the number of children, date of birth, gender, zip code, calls made, and account numbers can still be utilized to identify potential subjects. To stop these types of breaches, data mining must employ more stringent and highly effective privacy preservation procedures [26,27].

Here we are discussing the importance of privacy-preserving data mining techniques in preventing internet phishing and maintaining secure online information exchange. The Privacy Preserving Data Mining (PPDM) algorithm is introduced as an effective method for protecting sensitive information through altering and obscuring data.

Aims and objectives

1. To identify the extent of works and their limitations for cloud security

2. To design and develop the solution for cloud security, and to evaluate the model.

Questions of the study

1. Whether the formulated process can be helpful in the daily activities of the user.

2. Whether the designed process can maintain privacy policies and avoid intrusion of cloud computing.

Significance of the study

The research work's overview has been given. Centralized computing systems are now widely used as a result of the World Wide Web's explosive growth. In order to provide on-demand access to the content kept on the centralized network platforms, these solutions are essential. The term "cloud computing" generally refers to this strategy. At every point of the data lifecycle, data security is a requirement. Similarly, to this, strong security methods are needed for the data stages when the data is provided for use at sharing levels and stored in rest positions. Data remanence, which offers a physical representation of the data after the deletion procedure has been completed, is a frequent problem that has been raised from the data storage over the cloud. In order to do this, data encryption has been recognized as an appropriate approach that can aid in data protection. Along with these mentioned challenges, various aspects of the data kept in the cloud have flaws that affect its availability, confidentiality, and integrity. The common method by which the assurance of the stored data can be offered for defining the data transit has been highlighted as the integrity of the data. In this context, the term "data availability" refers to the reporting of pledge data.

2. Literature Review

Particularly as a result of the use of social networking sites like Twitter, Facebook, etc., data is dramatically growing every day. Because it costs more to maintain resources, it is extremely hard for data owners to preserve and manage such a large amount of data [15,16]. Data owners use cloud resources to reduce costs in order to solve this problem. However, the cloud might be curious as a 3rd party, which could result in the revelation of personal information. Due to this, researchers developed a technology known as privacy-preserving data mining that enables users of cloud storage to maintain their anonymity (PPDM). Private data is kept private even after mining since PPDM protects data owners' privacy in the cloud [28].

In a paper, they provide an improved defense mechanism (OpenFlowSIA) based on support vector machines and our suggested algorithm, the idle-timeout Adjustment, to guard Software-Defined Networks against flooding attacks (Distributed Denial-of-Service) (IA). In addition to effectively applying the IA algorithm and coherent policies to protect networks from resource exhaustion brought on by flooding attacks, their methodology also makes use of SVM classification advantages such as high accuracy and quick processing time. This is especially true for the SDN controller and OpenFlow switches. The OpenFlowSIA scheme demonstrates through extensive trials that it might be a creative approach to defend and conserve network resources from flooding attacks in Software-Defined Networks [29].

Recent technology developments have fueled the cloud's growth and success. Since it offers affordable structures that facilitate data transport, storage, and intense computing, this new paradigm is piquing growing interest. However, due to both the loss of data management and the impersonal nature of clouds, these prospective storage systems also come with a number of difficult design problems [17,18]. A recent poll was done with the aim of providing a unified perspective on client

concerns regarding data security and privacy in cloud storage environments. The protection of outsourced data in cloud infrastructures is addressed by a survey that offers a critical comparative analysis of cryptographic defense methods and, beyond this, discusses future research directions and technological trends [30].

Data mining aims to draw out meaningful information from massive sources of various data. However, the data becomes subject to processing during the data mining process, whether on purpose or accidentally. A novel idea in the field of data mining, privacy preservation places the security of user data mining as its top priority. It guarantees that sensitive data privacy will be maintained even after being mined by numerous parties. Data distortion, clustering, intersection, data distribution, and other known techniques for privacy-preserving data mining are used [31].

As a result of the Internet and social media, data and information are now more accessible and readily available than ever before. The data mining process is used to search this enormous data set and find undiscovered relevant data patterns and forecasts. Data mining enables the meaningful connection of unrelated data, the analysis of the data, and the representation of the results in the form of practical data patterns and predictions that aid in and forecast future behavior. Data mining has the potential to violate sensitive and private information. If some of the data leaks and identifies a person whose personal information was used in the data mining, then individual privacy is at risk. There is a variety of privacy-preserving data mining (PPDM) methodologies and procedures that are designed to protect sensitive data and privacy while still producing reliable data mining findings. Data protection strategies are incorporated into PPDM techniques and processes when data mining is being done [32].

3. Methods

New Model Mining

A new model mining proposed the primary way that this methodology operates is that the problem statement must be defined from a literature review. It covers the methods of processing in which the presentation of the plan must be created by taking into account the limitations of previous investigations. Based on these restrictions, the steps of the study methodology have been taken into account to define the extraction process. These steps have been illustrated as follows for clarity's sake:

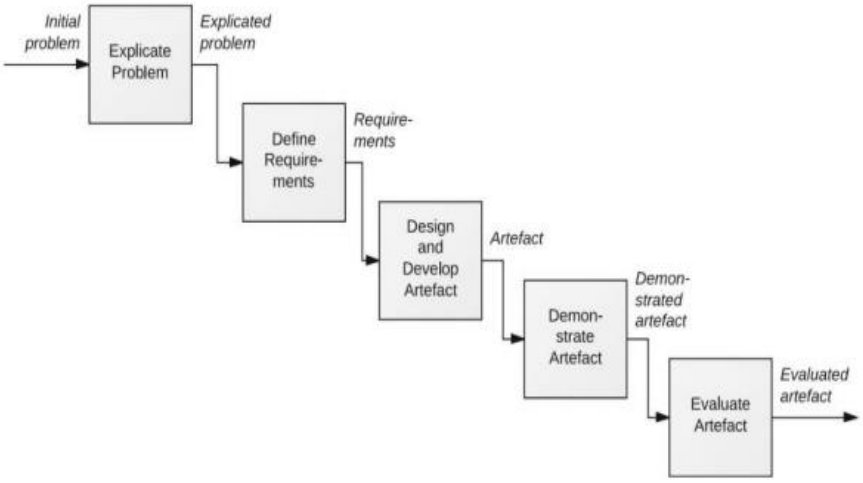

Fig 1: steps followed in this research methodology

The approaches that have been used to illustrate the development of the strategy for defining the limitation in existing models are presented in the identification of the problem statement. During the study, it was taken into account to identify the methods for dealing with DDoS attacks and preventing security concerns. The following chart shows the attacks over time (Figure 2).

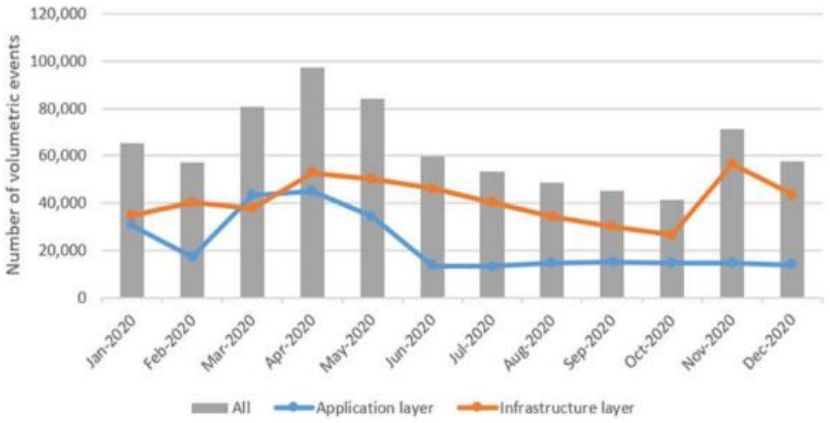

Fig 2: Trends in DDoS attacks over time.

The primary goal is specified in terms of illuminating the design of the model that has been chosen for development, in accordance with this step of the research technique. It comprises the procedures followed when creating the model, as well as the methods of processing that are used to illustrate the implementation strategy. Based on this stage, the choice of data mining methods has been taken into account for reporting the method to deal with DDoS attacks.

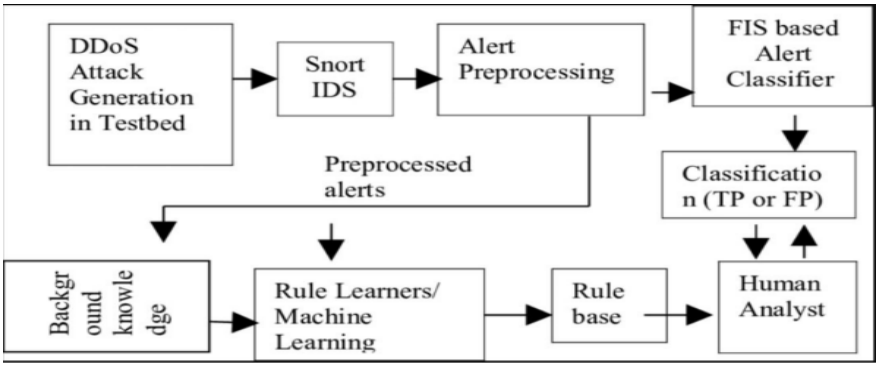

Fig 3: Flowchart illustrating the implementation strategy of Solution process for DDoS attacks

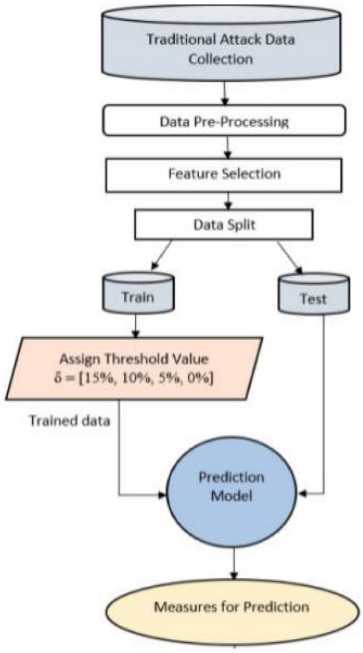

Fig 4: steps in methodology

Analysis of Meraz data

The Meraz dataset consists of several text data types, such as news articles, forum postings, and comments from social media. There are examples in various languages, including Spanish and French, along with the text data, which is mainly in English. The collection is well-annotated, and many of the text instances have labels. Machine learning models can be trained using these labels to perform tasks like sentiment analysis and text classification.

Table 1: Meraz dataset used

Attribute	**Description**
DataLabel	The label of the observation
EncryptionStatus	The status for the encryption data
FeatureSize	The Size of the model features
AlignmentFile	Model feature file alignment
FeatureModel	The size of models
FactorSize	The size of the packets
Checksum	Feature values of checksum factors

The bias in the Meraz'18 dataset can be found and measured using a variety of ways. One method is to analyze the text using a language model that has already been trained, and then look for patterns in the data that might point to bias. Another strategy is to have human annotators mark the text for bias before using the labeled data to train a machine-learning model to recognize bias in fresh content.

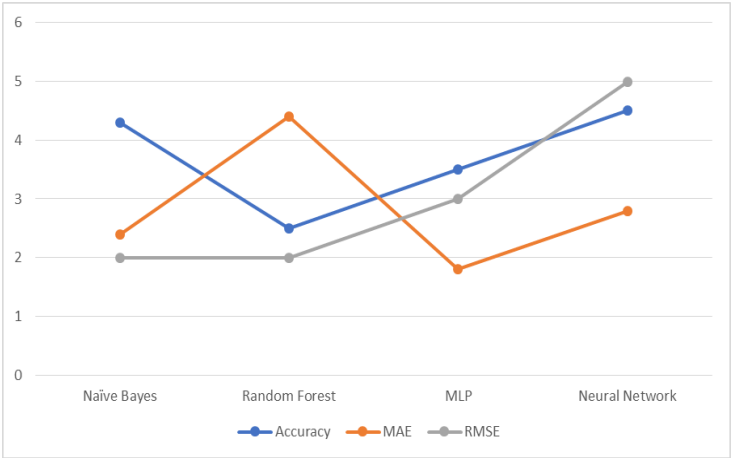

Fig 5: Result of analysis of accuracy in this study

4. Results

Organizations are increasingly using cloud computing to store, process, and analyze data, but because there is a higher danger of malware assaults, they must be able to gather and analyze significant amounts of data on malware.

We performed 6 experiments on the dataset and successfully, we get the following results:

Table 2**:** Comparison of the results of the algorithms

Experiments	**Accuracy**	**RMSE**
SVM	0.90	0.10
Naïve Bayes	0.75	0.30
Random Forest	0.80	0.20
Decision Tree	0.80	0.28
Logistic Regression	0.70	0.35
Neural Network	0.85	0.15

Based on this table, the best method would be SVM with the highest accuracy of 0.90 and the lowest RMSE of 0.10.

Deploying sensors or agents on cloud-based systems and apps is often the first step in the data collection process for malware investigation in the cloud. Network-based intrusion detection systems (NIDS) or intrusion prevention systems are a popular way to collect data (IPS). Utilizing endpoint security solutions, such as endpoint detection and response (EDR) or endpoint protection platforms (EPP), which identify malware at the endpoint level, is an additional strategy.

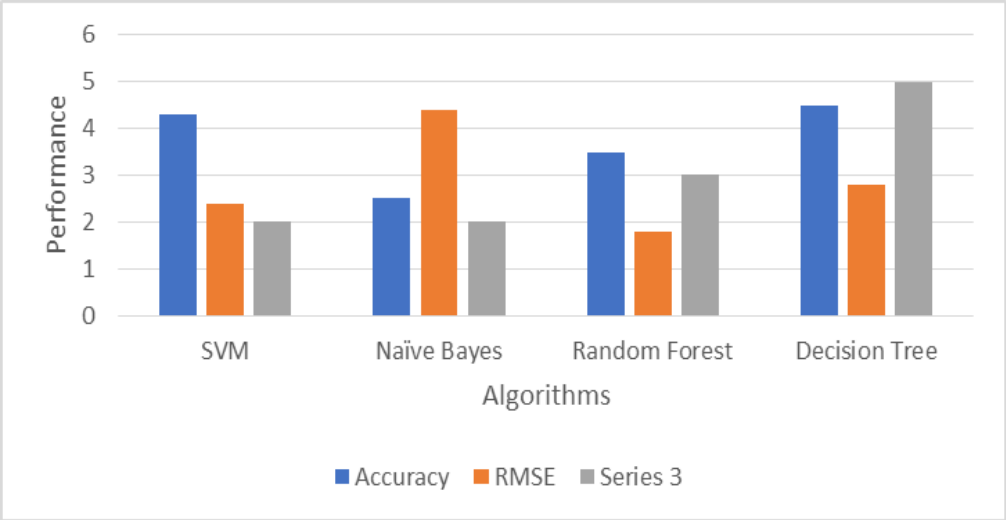

Fig 6**:** Various algorithms used in the study for analyzing intrusion features and their respective performance

In addition to these techniques, information can also be gathered directly from cloud-based applications and systems like web servers and databases. After the data has been gathered, it is processed and analyzed to find potential malware risks. This process incorporates machine learning, which aids in finding new and undiscovered malware. Organizations can take appropriate action, such as isolating infected systems and patching vulnerabilities, to stop the spread of the malware and lessen the impact of the assault when the data has been reviewed and the scale of the incident is established.

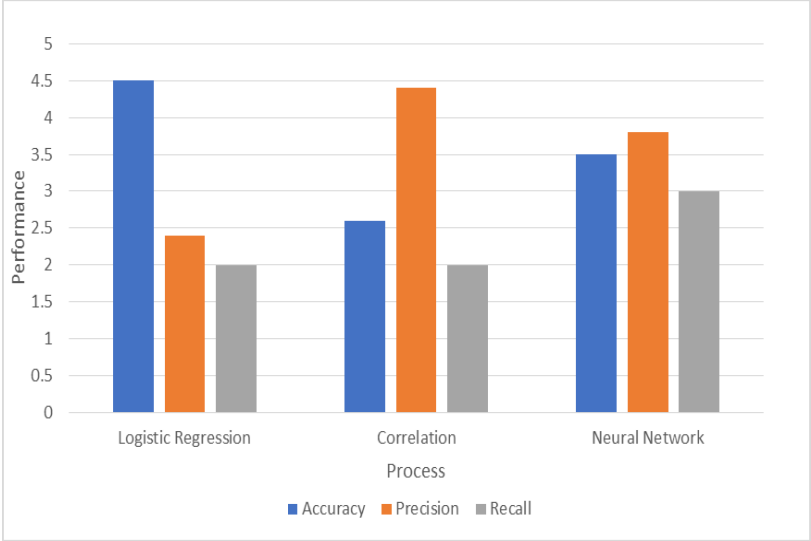

Fig 7**:** Analysis of designed models used in this study with their respective performance

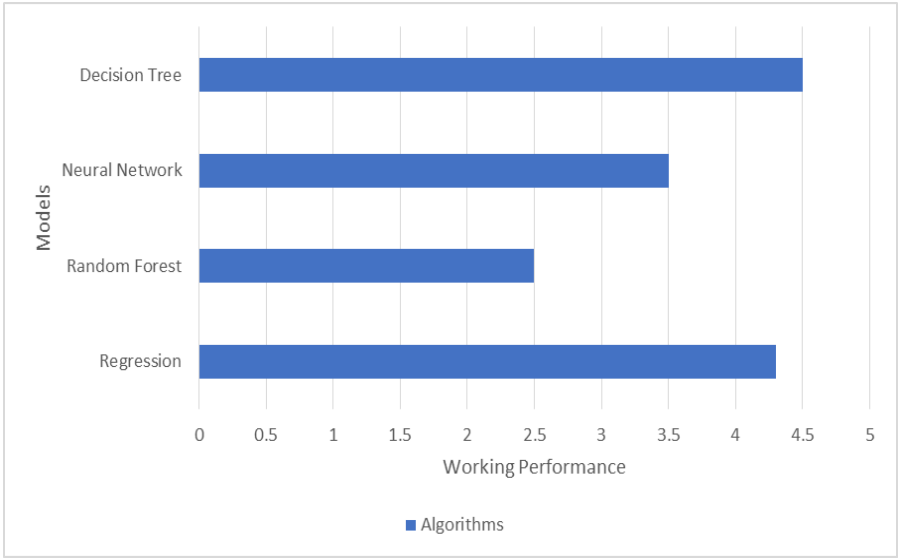

Fig 8**:** Models used in this study for detecting malware and their respective performance

Having access to high-quality datasets that can be utilized to train and test detection algorithms is important in order to successfully prevent and detect malware in cloud environments. Using publicly accessible datasets as a starting point in the development of a dataset for malware detection in the cloud is the first strategy. These datasets offer a significant quantity of data that can be utilized to train and test detection algorithms.

Using proprietary datasets gathered by the companies that run cloud environments is another method for creating a dataset for malware detection in the cloud.

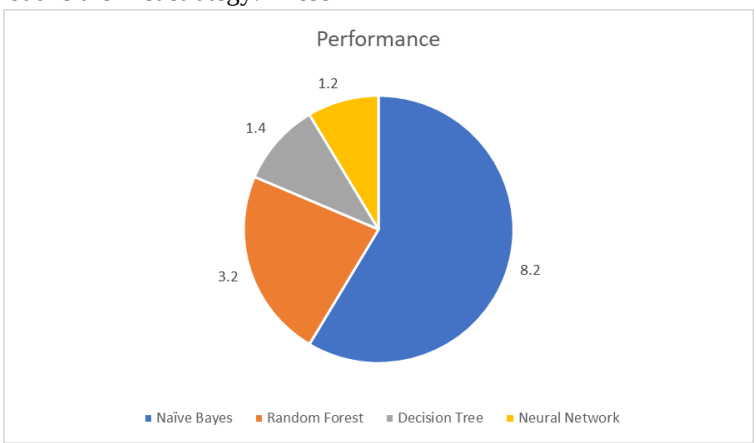

Fig9**: I**dentified attacks from different countries.

Utilizing synthetic datasets is a third method for creating a dataset for malware detection in the cloud. The use of active learning, which depends on detection algorithms, is a fourth method for creating a dataset for malware detection in the cloud. As a result, a variety of techniques can be utilized to create datasets for malware detection in the cloud.

This type of extracting the data is crucial in analyzing network security, however many limitations are present that impact the data collection.

One of the major limitations is the availability of data although the internet is a vast source of information everything is not relevant and some are not available as they are hidden or protected by encryption. Second is the quality of data. Third is the complexity of extracting the data as it is time-consuming and requires special tools, methods, and techniques. The fourth is privacy concern and the fifth is the scalability of data extraction. In order to overcome these limitations security analysts should select data carefully and maintain high knowledge of the field and also the organizations should follow new trends.

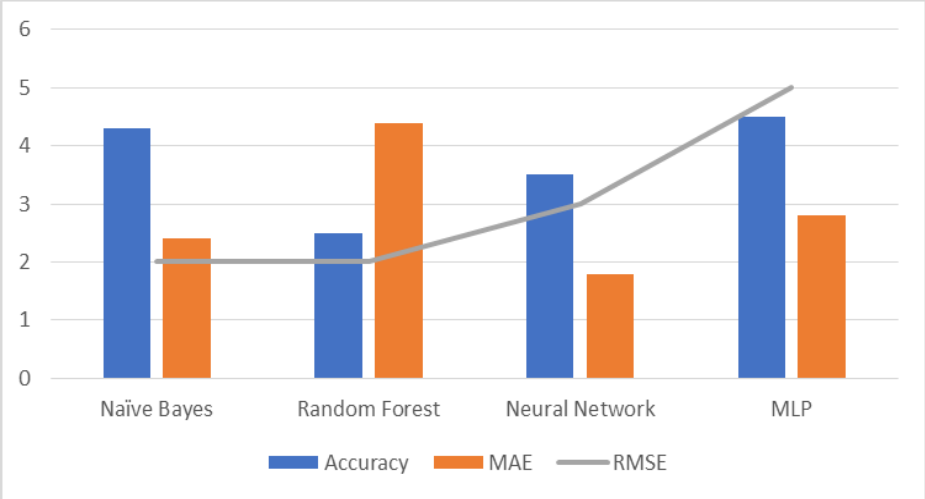

Fig 10**:** Implementation of machine learning models.

A combination of tools, techniques, and best practices must be used to implement the analysis of data using a dataset for malware analysis in cloud computing settings in order to successfully identify and stop malware assaults. The steps in implementing data analysis are data collection, data pre-processing, data visualization and exploration, feature engineering, model testing and training, model deployment, model maintenance and monitoring, and incident response.

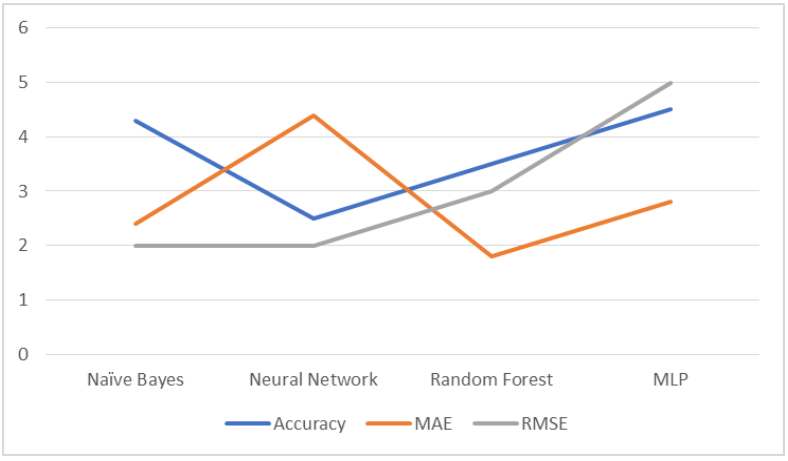

Fig 11**:** results of malware analysis vertices in this study

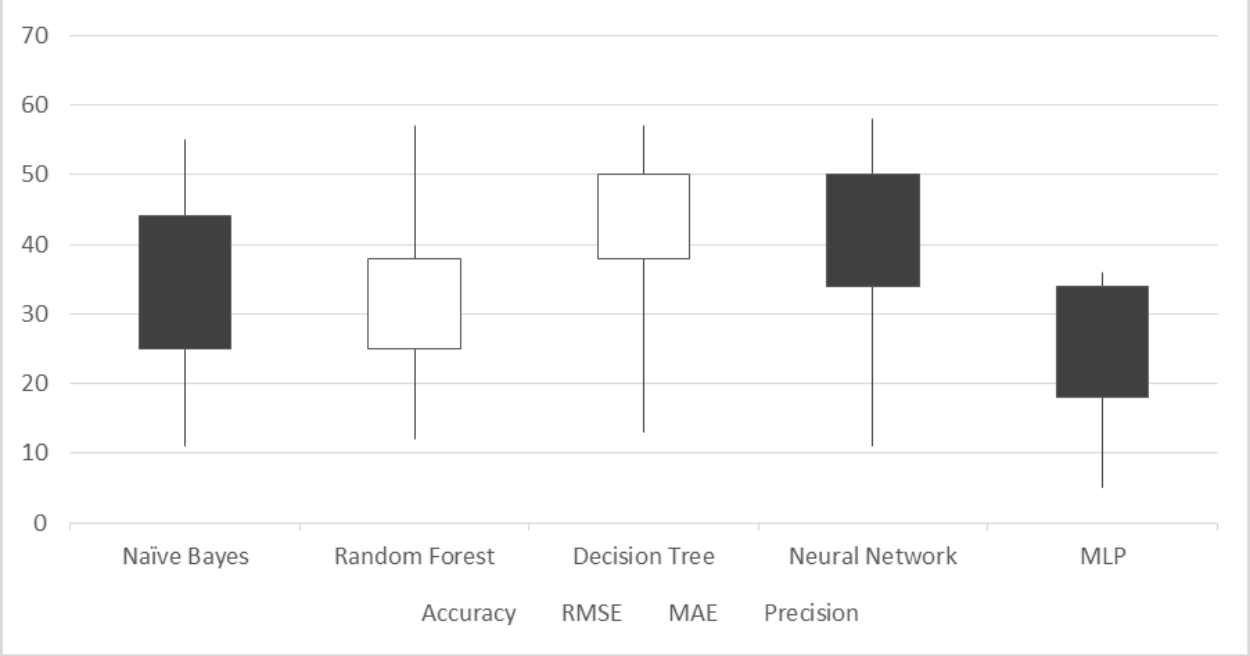

Fig 12**:** Results of malware analysis in this study

5. Discussion

The development of sophisticated cyberattacks that outperform conventional security defenses and have terrible repercussions has been driven by revolutionary advancements in network technologies. In order to build a strong line of defense against cyberattacks, intrusion detection systems (IDS) are now viewed as key components in network security infrastructures. High dimensionality's curse and class imbalance, which tend to lengthen detection times and reduce IDS efficiency, are the main difficulties facing IDS. Using four distinct types of classifiers—K-nearest neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Deep Belief Network (DBN)—the selected features of classification accuracy of an IDS model are examined. The analysis's findings indicate that RF is the best classifier to combine with any of these four feature assessment metrics in order to get a higher detection accuracy than other classifiers. We advise using consistency measures for constructing an effective IDS in terms of DR and FAR based on the statistical findings. In this study we used random first, SVM, decision tree, and naive bayes are used for analyzing accuracy [33].

Malicious software is referred to as malware. It is computer code that was created with malicious intent. A malware detector is a system that utilizes the call graph technique of the Application Programming Interface (API) as well as other techniques to try and identify malware. Because of its computational complexity, matching the API call graph using a graph-matching method has an NP-complete problem and is slow. This paper suggests an API call graph-based malware detection solution. A data-dependent API call graph is used to represent each malware sample. The input sample is first converted into a condensed data dependant graph, and then a database of malware API call graph samples is compared to them using a graph-matching algorithm. The Longest Common Subsequence (LCS) algorithm, which is applied to simplified networks, is the basis of the graph matching algorithm. By choosing paths in the API call graph with the same edge label, such an approach lowers the computational complexity. The suggested malware detection method has a 9.38% detection rate and a 0% false positive rate, according to experimental results on 85 samples [34].

Predictive maintenance is a condition-based maintenance approach (CBM) that only performs maintenance when it is necessary, preventing breakdowns or pointless preventive actions. Advanced monitoring and diagnosis solutions using machine learning (ML) have grown more and more appealing. Implementing ML-based PdM is a challenging and expensive procedure, particularly for those businesses that frequently lack the essential talent, resources, and labor. So, in order to determine when ML-based PdM is the best maintenance technique, a cost-oriented analysis is necessary. However, no prior research has taken both costs into account in the economic evaluation of PdM. The implementation of this strategy involves investment costs in IT technologies in addition to costs incurred from traditional maintenance activities depending on the performance of the ML model classifier [35]

6. Conclusion

During this study, we identified the main obstacle of cloud computing as the incorporation of security protocols to protect the data stored in the cloud. To address this, we developed a process that enhances the user's daily activities by enabling them to store and retrieve data from a centralized platform. The process also takes into account the latest trends in the field of cloud computing to ensure the security of the data. We implemented the proposed model and achieved results and accuracy when we applied the svm algorithm, the accuracy was 90% and the root mean squared error was 10%.. Our model employed various technologies, including DDoS assault prevention mechanisms, to ensure secure data storage. Overall, this study highlights the importance of maintaining privacy policies and offers a comprehensive solution for the security of data stored in the cloud.

Acknowledgment

The authors would like to thank University of Kufa and University of mustansiriyah for its support in this work.

References

[1] Aggarwal G, Bawa M, Ganesan P (2005) Two can keep a secret: a distributed architecture for secure database services. CIDR

[2] Agrawal R, Srikant R (2000) Privacy-preserving data mining. In: Proceedings of the 2000 ACM SIGMOD international conference on management of data—SIGMOD '00, vol 29, no 2. pp 439–450. http://doi.org/10.1145/342009.335438

[3] Aggarwal CC, Yu PS (2008) A general survey of privacy-preserving data mining models and algorithms. In: Privacy preserving data mining, Chap 2. Springer, New York, pp 11–52. http://doi.org/10.1007/978-0-387-48533

[4] Arunadevi M, Anuradha R. Privacy preserving outsourcing for frequent itemset mining. *Int J Innov Res Comp Commun Eng.* 2014;2(1):3867–3873.

[5] Baotou T (2010) Research on privacy preserving classification data mining based on random perturbation Xiaolin Zhang Hongjing Bi. 1–6

[6] Belwal R, Varshney J, Khan S (2013) Hiding sensitive association rules efficiently by introducing new variable hiding counter. In: IEEE international conference on service operations and logistics, and informatics, 2008, IEEE/SOLI 2008, vol 1, pp 130–134. doi:10.1109/SOLI.2008.4686377

[7] Chan J, Keng J (2013) Privacy protection in outsourced association rule mining using distributed servers and its privacy notions, pp 1–5

[8] Ciriani V, Vimercati SDC, Foresti S, Samarati P. *Privacy-preserving data mining.* New York: Springer; 2008. k-anonymous data mining: a survey; pp. 105–136.

[9] Dehkordi MNM, Badie K, Zadeh AKA. A novel method for privacy preserving in association rule mining based on genetic algorithms. *J Softw.* 2009;4(6):555–562. doi: 10.4304/jsw.4.6.555-562.

[10] Deivanai P, Nayahi J, Kavitha V (2011) A hybrid data anonymization integrated with suppression for preserving privacy in mining multi party data. In: IEEE international conference on recent trends in information technology (ICRTIT)

[11] Dev H, Sen T, Basak M, Ali ME (2012) An approach to protect the privacy of cloud data from data mining based attacks. In: IEEE 2012 SC companion high performance computing, networking, storage and analysis (SCC)

[12] Ding X, Yu Q, Li J, Liu J, Jin H (2013) Distributed anonymization for multiple data providers in a cloud system. In: Database systems for advanced applications. Springer, Berlin, Heidelberg

[13] Domadiya NH, Rao UP (2013) Hiding sensitive association rules to maintain privacy and data quality in database. In: IEEE 3rd international on advance computing conference (IACC), pp 1306–1310

[14] Dong R, Kresman R (2009) Indirect disclosures in data mining. In: Fourth international conference on frontier of computer science and technology, FCST'09

[15] Friedman A, Wolff R, Schuster A. Providing k-anonymity in data mining. *VLDB J.* 2008;17(4):789–804. doi: 10.1007/s00778-006-0039-5.

[16] Giannotti F, Lakshmanan LVS, Monreale A, Pedreschi D, Wang H. Privacy-preserving mining of association rules from outsourced transaction databases. *IEEE Syst J.* 2013;7(3):385–395. doi: 10.1109/JSYST.2012.2221854.

[17] Gkoulalas-Divanis A, Verykios VS. Exact knowledge hiding through database extension. *IEEE Trans Knowl Data Eng.* 2009;21(5):699–713. doi: 10.1109/TKDE.2008.199.

[18] Harnsamut N, Natwichai J (2008) A novel heuristic algorithm for privacy preserving of associative classification. In: PRICAI 2008: trends in artificial intelligence. Springer, Berlin, Heidelberg, pp 273–283

[19] He Y, Siddharth B, Jeffrey FN (2011) Preventing equivalence attacks in updated, anonymized data. In: IEEE 27th international conference on data engineering (ICDE)

[20] Ibrahim A, Jin H, Yassin AA, Zou D (2012) Towards privacy preserving mining over distributed cloud databases. In: IEEE second international conference on cloud and green computing (CGC)

[21] Inan A, Saygin Y (2010) Privacy preserving spatio-temporal clustering on horizontally partitioned data. In: Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6202 LNAI. pp 187–198. http://doi.org/10.1007/978-3-642-16392-0_11

[22] Islam MZ, Brankovic L. Privacy preserving data mining: a noise addition framework using a novel clustering technique. *Knowl Based Syst.* 2011;24(8):1214–1223. doi: 10.1016/j.knosys.2011.05.011.

[23] Jain YYK, Yadav VKVVK, Panday GGS. An efficient association rule hiding algorithm for privacy preserving data mining. *Int J Comp Sci Eng.* 2011;3(7):2792–2798.

[24] Kamakshi P (2012) Automatic detection of sensitive attribute in PPDM. In: IEEE international conference on computational intelligence & computing research (ICCIC)

[25] Kamakshi P, Babu AV (2010) Preserving privacy and sharing the data in distributed environment using cryptographic technique on perturbed data 2(4)

[26] Kantarcıoglu M, Vaidya J (2003) Privacy preserving naive bayes classifier for horizontally partitioned data. In: IEEE ICDM workshop on privacy preserving data mining, pp 3–9. http://www.cis.syr.edu/~wedu/ppdm2003/papers/1.pdf

[27] Karim R, Rashid M, Jeong B, Choi H (2012) In: Transactional Databases. pp 303–319

[28] Chandrakar, I., & Hulipalled, V. R. (2020). Techniques for Preserving Privacy in Data Mining for Cloud Storage: A Survey. *Advances in Intelligent Systems and Computing*, 452–461. https://doi.org/10.1007/978-981-15-6353-9_42

[29] Phan, Trung & Toan, Truong & Van Tuyen, Dang & Huong, Truong & Thanh, Nguyen. (2016). OpenFlowSIA: An optimized protection scheme for software-defined networks from flooding attacks. 13-18. 10.1109/CCE.2016.7562606.

[30] Kaaniche, N., & Laurent, M. (2017). Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms. *Computer Communications*, *111*, 120-141. https://doi.org/10.1016/j.comcom.2017.07.006

[31] Purohit, R., & Bhargava, D. (2017). An illustration to secured way of data mining using privacy preserving data mining. *Journal of Statistics and Management Systems*, *20*(4), 637-645.

[32] Kreso, I., Kapo, A., & Turulja, L. (2021). Data mining privacy preserving: Research agenda. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(1), e1392.

[33] Binbusayyis, A., & Vaiyapuri, T. (2020). Comprehensive analysis and recommendation of feature evaluation measures for intrusion detection. *Heliyon*, *6*(7). https://doi.org/10.1016/j.heliyon.2020.e04262

[34] Elhadi, Ammar & Maarof, Mohd & Barry, Bazara. (2013). Improving the Detection of Malware Behaviour Using Simplified Data Dependent API Call Graph. International Journal of Security and Its Applications. 7. 29-42. 10.14257/ijsia.2013.7.5.03.

[35] Florian, E., Sgarbossa, F., & Zennaro, I. (2021). Machine learning-based predictive maintenance: A cost-oriented model for implementation. *International Journal of Production Economics*, *236*, 108114. https://doi.org/10.1016/j.ijpe.2021.108114