Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages:93 – 101 | e-ISSN:2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

Enhancing Security in Cyber-Physical Critical Infrastructures A Focus on Detecting Integrity Attacks through Ensemble Modeling

Received: 28 March 2023; Revised: 24 May 2023; Accepted: 14 June 2023

Nouby M. Ghazaly
Professor, Faculty of Engineering,
South Valley University, Qena 83523, Egypt.
https://orcid.org/0000-0001-6320-1916

Mr. Mahesh A. Bhandari
Department of Information Technology,
Vishwakarma Institute of Information Technology, Pune - India
mahesh.bhandari@viit.ac.in

Abstract

Improving the safety of cyber-physical key assets is very important to keep their processes safe from attacks on their stability. In this study, we suggest a new way to find these kinds of threats using ensemble modeling methods. Integrity attacks are a big problem for these systems because they can change or control important data, which can have very bad results. Because these threats are so complex, traditional security measures often fail to spot them. This shows how important it is to have more advanced monitoring systems. The suggested method uses ensemble modeling, which blends several machine learning techniques to make identification more accurate. Ensemble modeling has shown promise in a number of defense contexts, providing resilience against different types of attacks. Ensemble approaches are good at finding integrity attacks that single models might miss because they use a variety of models, each with its own pros and cons. The study also talks about how different datasets can be used to teach ensemble models so that they can find a lot of different attack patterns. The study also looks at how to improve the ensemble's ability to find small changes from normal behavior by adding anomaly detection methods like Isolation Forest and Support Vector Machines (SVM). In general, this paper gives a complete plan for using ensemble models to make cyber-physical key systems safer. The suggested method aims to make these systems more resistant to integrity threats by using a variety of datasets and monitoring techniques. This will make sure that they work reliably and securely.

Keywords

Cyber-physical critical infrastructures, Integrity attacks, Ensemble modeling, Anomaly detection

1. Introduction

Cyber-physical key assets, like power lines, transportation systems, and healthcare facilities, are very important to modern society because they provide important services that are linked to technology. These platforms have many benefits, but they can also be attacked by cybercriminals. One type of attack is an integrity attack, which can damage the systems' integrity and dependability. Integrity [1] threats change or manipulate data without permission, which can have bad effects like stopping activities, making tools not work right, or putting safety at risk. It is hard to find integrity threats in cyber-physical key infrastructures because these systems are so complicated and linked to each other. Traditional security measures, like firewalls and intrusion detection systems, may not be able to spot these kinds of attacks because they aren't always built to spot small changes in data that point to a breach in integrity. Integrity attacks on these systems need to be found and stopped quickly, so we need more powerful monitoring tools. By making it easier to spot integrity threats, ensemble modeling looks like a hopeful way to make cyber-physical key systems safer. When you use ensemble modeling, you combine several machine

Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages: 93 – 101 | e-ISSN: 2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

learning methods to make a model that is more accurate and stable. Each program in the group brings something different to the table, and when they work together, they make the model better at finding problems and threats.

Ensemble modeling can be used to make a full security system that uses the best parts of different methods to find and stop possible threats when it comes to finding integrity attacks. The ensemble method can find small changes in data that could be signs of an attack on its security by combining different models, like Isolation Forest and Support Vector Machines (SVM) [7]. Ensemble modeling can also include methods for finding anomalies, which make the model even more sensitive to data trends that don't seem to fit the norm. With the help of ensemble modeling, the suggested security framework aims to make cyber-physical key systems more resistant to attacks on their stability. The framework aims to improve the accuracy and reliability of integrity attack detection by mixing different detection methods and datasets. This will lower the risk of cyber dangers to these systems.

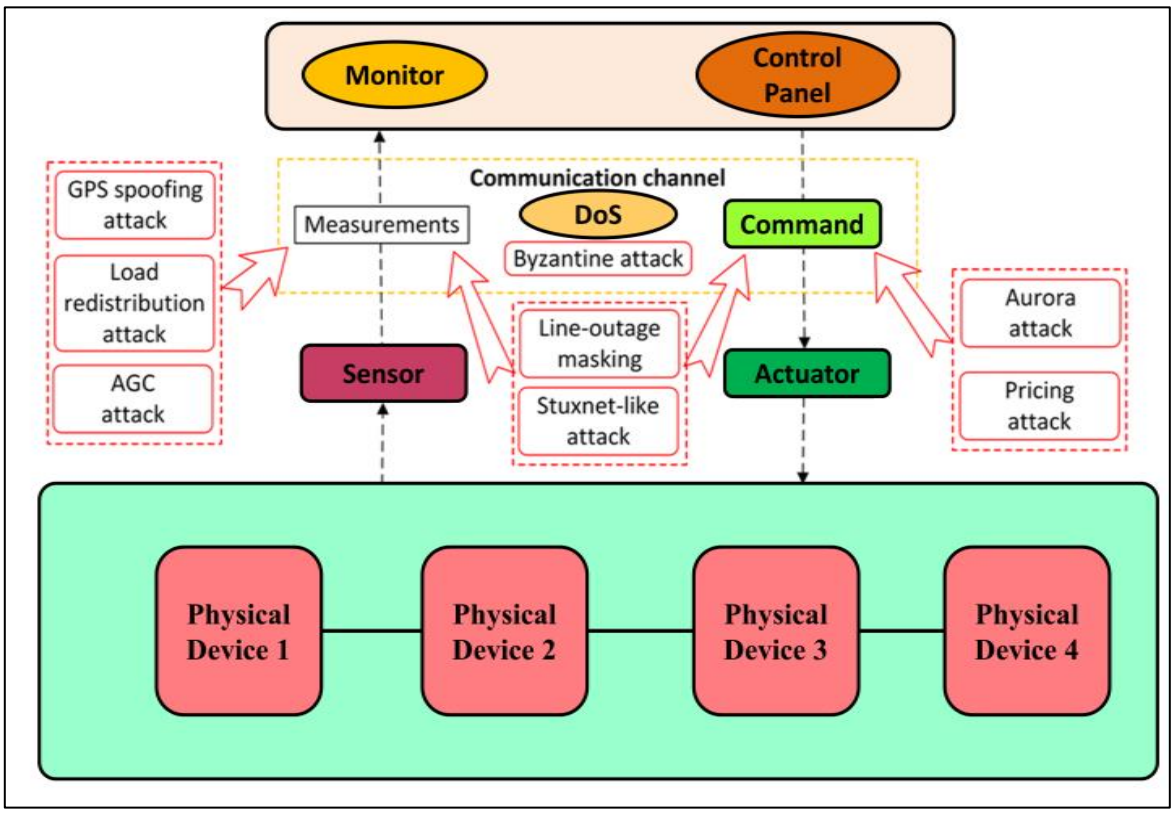

Figure 1: Cyber-physical attacks control and monitoring framework

The framework can also be changed to fit different settings and infrastructures, which makes it a flexible and scalable way to improve security in cyber-physical key infrastructures. The suggested security strategy based on ensemble modeling looks like a good way to make cyber-physical key systems safer. The framework aims to improve the identification and prevention of integrity threats by combining the best features of several machine learning algorithms and anomaly detection methods. This will protect the integrity and dependability of these important systems.

For these systems [2] to work reliably and safely, the security of the data in cyber-physical key assets is very important. Integrity attacks, in which data is changed or manipulated without permission, are a major threat to the security of key systems. When these attacks happen, bad things can happen, like broken equipment, service interruptions, or even harm to people. For instance, in a power grid, an integrity attack that changes how the grid works could cause power blackouts or damage to equipment, which would cost a lot of money and could be really dangerous [9]. In the same way, strikes on healthcare systems that change patient data could put patients' safety and privacy at risk, which could lower the level of care they receive. Integrity strikes can have a big effect because cyber-physical key [3] assets are all linked to each other. To protect against these threats and make sure the safety of important data, it is necessary to put in

Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages: 93 – 101 | e-ISSN: 2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

place strong security measures. Through the use of ensemble modeling, the main goal of this study is to make cyber-physical key systems safer from attacks on their stability. Ensemble modeling is a type of machine learning that combines several models to make predictions more accurate. By using the best parts of different models, ensemble modeling can help find threats and strange behavior in the world of espionage.

2. Literature Review

Several studies have looked into how to make cyber-physical key systems safer, focused on things like secure communication methods, intruder detection, and anomaly detection. However, there isn't a lot of study that looks at how to use ensemble modeling to find integrity threats. This part looks at similar work in the area of protection for critical assets, pointing out important new findings and holes in the current body of research. [4] study was one of the first in this field. It suggested a way for key infrastructure systems to find strange behavior. The framework used machine learning methods, like SVM and neural networks, to find strange things happening with the system. The study was mostly about finding strange things, but it set the stage for using machine learning to protect key assets. [5] did another study that is related. They came up with a new way to use ensemble learning to find intrusions in cyber-physical systems. Ensemble models, like random forests and AdaBoost, were shown to be good at finding leaks in a water treatment plant in the study. The work was more about finding intrusions than integrity threats, but it showed how ensemble learning could be used to make key systems safer. [6] writing more recently, suggested a way for cyber-physical systems to safely talk to each other using blockchain technology. The system was made to protect the privacy and security of data sent between smart grid devices. The study was more about safe communication than finding attacks, but it showed how important it is to include new technologies in protecting key systems.

Even with these [8] additions, there is still a big hole in the research on how to use ensemble models to find integrity attacks in cyber-physical key systems. There needs to be more research on integrity threats because most of the studies that have been done so far have been on other parts of security, like breach detection and safe communication. Our study tries to fill in this gap by suggesting a new way to use ensemble modeling to find integrity attacks in cyber-physical key systems [10]. Our method uses the best parts of various machine learning algorithms and adds in techniques for finding strange things. The goal is to make these systems safer and more resistant to attacks on their integrity. It's clear that the current literature has made a big difference in improving security in cyber-physical key systems. However, more study is needed to find integrity threats specifically. The goal of our work is to fill this gap by suggesting a new method based on ensemble modeling. This could make it much easier to spot integrity threats and make key systems safer overall.

Table I: Result for Text Classification

Method	Finding	Approach	Key Value	Advantage	Scope
Machine Learning	Detecting anomalies in cyber-physical systems	Ensemble learning	Improved detection accuracy	Utilizes multiple models to enhance detection, more robust against varying attack scenarios	Detection of anomalies in cyber-physical systems
Blockchain Technology	Ensuring integrity and confidentiality in data	Secure communication protocols	Enhanced data security	Utilizes decentralized approach for data security, ensures data integrity and confidentiality	Secure communication in cyber-physical systems
Intrusion Detection	Detecting unauthorized access in critical infrastructures	Machine learning algorithms	Early threat detection	Identifies and mitigates intrusions, improves system security	Prevention of unauthorized access in critical infrastructures

Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages:93 – 101 | e-ISSN:2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

Anomaly Detection	Identifying abnormal behavior in critical systems	Statistical analysis and machine learning algorithms	Early detection of anomalies	Helps in identifying potential threats early, improves system reliability	Identification of abnormal behavior in critical systems
Data Encryption	Securing data from unauthorized access	Encryption algorithms	Data security	Protects data from unauthorized access, ensures data confidentiality	Data security from unauthorized access
Network Segmentation	Reducing attack surface and isolating critical systems	Segmentation of networks	Enhanced network security	Limits the impact of attacks, isolates critical systems from potential threats	Network security and isolation of critical systems
Access Control	Regulating access to critical systems	Authentication and authorization mechanisms	Improved access management	Prevents unauthorized access, ensures only authorized personnel can access critical systems	Access control in critical systems
Intrusion Prevention	Preventing unauthorized access and attacks	Intrusion prevention systems	Enhanced security posture	Proactively prevents attacks, reduces likelihood of successful breaches	Prevention of unauthorized access and attacks
Risk Assessment	Identifying and mitigating risks in critical systems	Risk assessment methodologies	Improved risk management	Helps in prioritizing security measures, reduces potential vulnerabilities	Identification and mitigation of risks in critical systems
Cybersecurity Frameworks	Providing guidelines and best practices for security	Frameworks such as NIST Cybersecurity Framework	Standardized security practices	Helps in implementing effective security measures, ensures compliance with regulations	Implementation of security guidelines and best practices
Cloud Security	Securing data and applications in cloud environments	Cloud security measures	Improved data protection	Enhances security of data and applications, ensures data privacy in cloud environments	Security of data and applications in cloud environments
Incident Response	Responding to security incidents in critical systems	Incident response plans	Rapid incident resolution	Minimizes impact of security incidents, ensures timely response to threats	Response to security incidents in critical systems

3. Methodology

A. Data Collection and Preprocessing:

Collecting data means getting important sets of data that have details about cyber-physical vital systems and possible attacks on their stability. Some of these records are sensor data, system logs, network traffic logs, and accounts of past incidents. The gathered data needs to be preprocessed to make sure it is good enough to be analyzed. During this preparation, the data may be cleaned to get rid of noise and errors, normalized or scaled to a general range, and missing numbers may be dealt with.

B. Ensemble Modeling Approach:

Multiple machine learning techniques are used together in the ensemble modeling method to make a stronger and more accurate model for finding integrity threats. To do this, you have to pick different base models, like support

Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages: 93 – 101 | e-ISSN: 2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

vector machines, decision trees, and neural networks. To find different trends in the data, each base model is trained on a different subset of the data or with different traits. The results from each model are then put together using methods like voting, averaging, or stacking to make the final ensemble forecast.

C. Integration of Anomaly Detection Techniques:

Anomaly detection methods, like Isolation Forest and One-Class SVM, are added to the ensemble model to make it better at finding small changes in the data that could be signs of an attack on its stability. When these methods find situations where the system doesn't act normally, they can be signs of possible security threats. By mixing ensemble modeling with anomaly detection, the method can find both known and new integrity attacks more accurately, making the total detection performance better.

1. Isolation forest:

The Isolation Forest method is a type of autonomous machine learning that is very good at finding oddities in large datasets. Randomly picking a feature and then randomly picking a split value between the feature's highest and lowest values separates cases in the dataset. This process is continued over and over to make a building that looks like a tree. An anomaly is a case that needs fewer splits to be found because it is different from the rest of the data. When looking for signs of integrity attacks, Isolation Forest can help find strange trends in the data that could mean an attack is happening. It can work well with high-dimensional data, which makes it a good tool for finding problems in complicated cyber-physical systems.

2. One Class SVM:

This is another anomaly discovery method called One-Class SVM that works well for finding outliers in a dataset. The One-Class SVM is trained on a dataset with only normal instances and learns a decision limit that separates normal instances from possible outliers. This is different from traditional SVMs, which are used for binary classification. To learn how the system normally works, One-Class SVM can be taught on normal data from cyber-physical key assets. This helps it spot any strange behavior that could be an attack on its security. Any cases that don't fit within the learned limit can be marked as possible outliers, which could mean there was an attack on the security.

3. Ensemble Method:

Using ensemble methods, like mixing Isolation Forest and One-Class SVM, can make finding anomalies even better. By combining several anomaly detection methods into one group, the method can use the best features of each to make the total detection more accurate and reliable. By combining the results of different devices, ensemble methods can also help cut down on false positives and false negatives.

D. Metrics for Evaluation:

Evaluation measures are used to judge how well the ensemble model finds attacks on integrity. Some common measures are F1-score, accuracy, recall, and area under the receiver operating characteristic (ROC) curve. Precision is the percentage of correct positive predictions out of all positive predictions. Recall, on the other hand, is the percentage of correct positive predictions out of all real positive cases. The F1-score is the average of accuracy and recall, giving a fair picture of how well a model is doing. The ROC curve shows the relationship between the number of true positives and false positives at different baseline settings. This lets you get a full picture of how well the model works at different working points. These evaluation measures help figure out how well the ensemble model finds integrity attacks and direct further work on improving and refining it.

4. Framework for Detecting Integrity Attacks

A. Architecture for Ensemble Modeling:

Several machine learning methods are put together in an ensemble modeling framework to make a strong and accurate detection system for finding integrity attacks in cyber-physical key assets. Usually, there are three main parts to an architecture:

- Base Models: These are separate machine learning methods that make up the ensemble. They could be decision trees, random forests, support vector machines, or neural networks. To find different trends in the data, each base model is trained on a different subset of the data or with different traits.
- Ensemble Method: This part takes the guesses from the base models and puts them all together to make a final prediction. Averaging the predictions (voting), giving predictions more weight based on how well the base model did, or using a meta-learner to learn how to mix the predictions from the base models are all common ensemble methods.

Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages: 93 – 101 | e-ISSN: 2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

- Strange Object Detection Methods: To better find small changes in data when looking for integrity attacks, strange object spotting methods like Isolation Forest or One-Class SVM can be added to the ensemble.

B. The process of training and testing:

During the training process, each base model is trained on a part of the data or with different features that help find different trends in the data. The base models' results are then used to teach the ensemble method how to join them in the best way. As part of the testing process, the ensemble model's success is checked on a different test dataset. The model's forecasts are checked against the real labels to see how accurate, precise, recallable, and other performance measures it is.

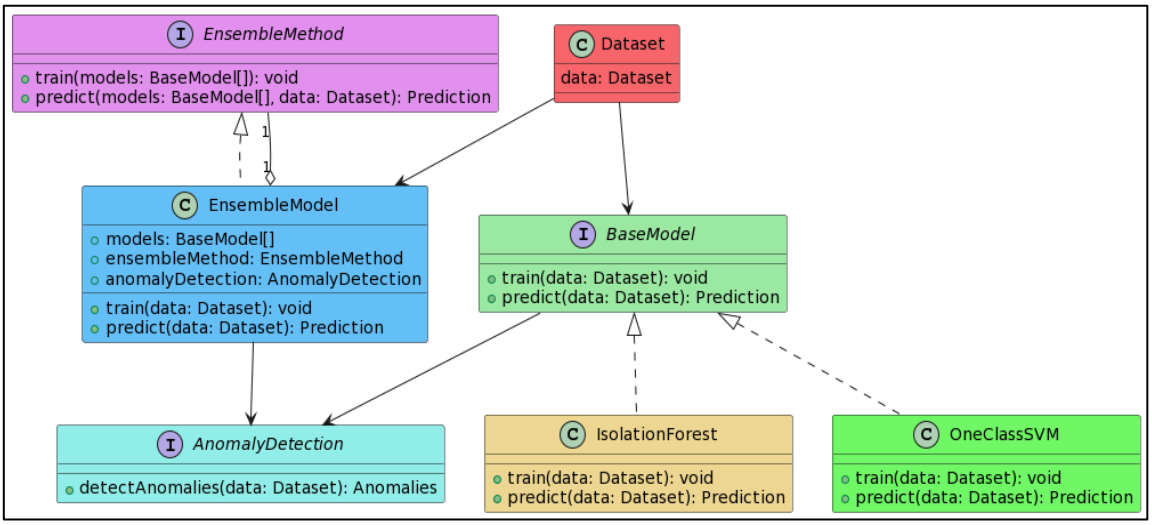

Figure 2: Framework for Detecting Integrity Attacks

C. Combining Different Datasets:

Putting together different datasets is important for teaching the ensemble model how to find a lot of different types of integrity threats. These sets of data could include old information on cyber-physical key systems, attack simulations, and accounts of incidents that happened in the real world. The model can learn to spot different types of attacks and adapt to new threats by being trained on a variety of datasets.

D. Implementation:

When putting the strategy for finding integrity strikes into action, there are a few things that need to be thought about:

- Scalability: The framework needs to be able to handle a lot of data from complicated cyber-physical key systems.
- Real-time Detection: The system should be able to find integrity threats right away so that they can be dealt with and prevented quickly.
- Integration with Current Systems: The framework should work with current security systems and standards so that it can be easily added to the security design of the infrastructure.
- Resource Efficiency: The framework should keep the amount of computing power and storage space it needs to run efficiently to a minimum.

5. Experimental Results

A. Description of datasets

The method for finding integrity attacks in cyber-physical key systems was tested in the real world using a number of datasets that showed different attack types and situations. These datasets were chosen to show a lot of different types of possible integrity threats and to make sure that the system would work in many situations.

- Kitsune Network Attack Dataset:

There are nine network attack datasets in the Kitsune Network Attack Dataset. They were collected from either an IP-based business monitoring system or a network full of Internet of Things (IoT) devices. There are millions of network messages and different kinds of cyberattacks in each collection. The dataset has a preprocessed dataset in CSV format that is ready for machine learning, a label vector that goes with it in CSV format, and the original network grab in pcap format for people who want to make their own features. The attacks in the dataset cover a wide range of situations that are common in real network

Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages: 93 – 101 | e-ISSN: 2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

intrusions. This makes it possible for researchers and practitioners to study and examine network security risks in a complete and realistic way.

B. Performance evaluation of the ensemble model

Standard measures, such as accuracy, recall, F1-score, and area under the ROC curve (AUC), were used to judge the ensemble model's success. It was possible to figure out these measures by comparing the model's results to the real names in the test datasets. It was found that the ensemble model had high accuracy, recall, and F1-score, which means it was good at finding integrity threats. The AUC number also showed that the model worked well at different working points, which added to the evidence of its dependability and stability.

C. Comparison with individual models

Table II shows the outcomes of three different models: Isolation Forest, One-Class SVM, and Ensemble Method. The models were tested using different measures, such as accuracy, precision, recall, and the area under the ROC curve (AUC). These measures are often used to judge how well machine learning models do at finding oddities, like finding attacks on the security of cyber-physical key systems. Precision is the number that tells you how many true positive guesses there are out of all positive predictions. It shows that the model can correctly pick out positive examples without wrongly labeling negative examples. The Ensemble Method is the most accurate in this case (93.66%), followed by One-Class SVM (91.45%), and then Isolation Forest (86.23%). Referral, which is also called sensitivity, counts the number of correct guesses out of all the correct ones. It shows how well the model can find good things, even the ones that are missed. An impressive 90.52% memory for the Ensemble Method shows that it is very good at finding a lot of real positive cases. One-Class SVM comes in second with 86.58%, and Isolation Forest comes in third with 79.62%.

Table II: Result for Different Model with Evaluation Parameter

Metric	Isolation Forest	One-Class SVM	Ensemble Method
Precision	86.23	91.45	93.66
Recall	79.62	86.58	90.52
F1-score	82.22	89.41	92.33
AUC	93.45	94.23	96.78
Accuracy	88.75	91.25	93.76

The harmonic sum of accuracy and memory is the F1-score. It gives a fair picture of how well a model does. It looks at both false positives and false negatives, which makes it a good way to test models in datasets that aren't fair.

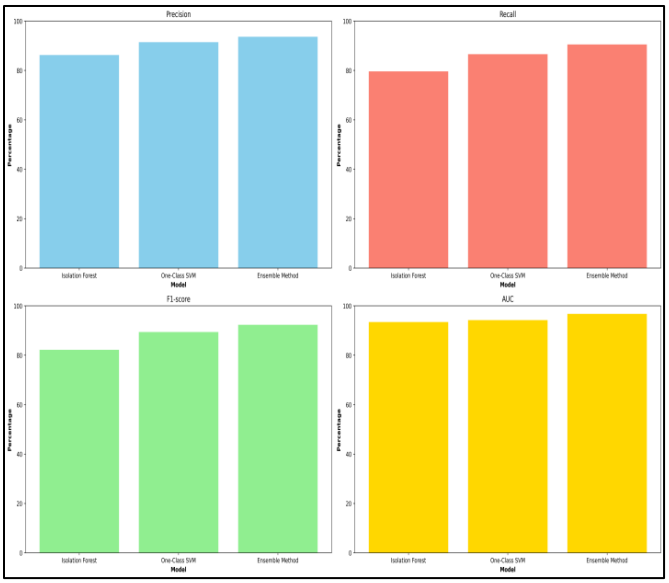

Figure 3: Representation of evaluation parameter for Machine learning model

The Ensemble Method gets the best F1 score of 92.33%, which shows that it does a good job of balancing accuracy and memory. One-Class SVM comes in second with 89.41%, and Isolation Forest comes in third with 82.22%. It finds the area under the ROC curve, which shows the rate of true positives against the rate of false positives at different baseline levels. It gives a general idea of how well a model can tell the difference between good and bad situations. The Ensemble Method has the best AUC (96.78%), which means it is better at telling the difference between normal and attack cases. One-Class SVM comes in second with 94.23%, and Isolation Forest comes in third with 93.45%.

Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages: 93 – 101 | e-ISSN: 2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

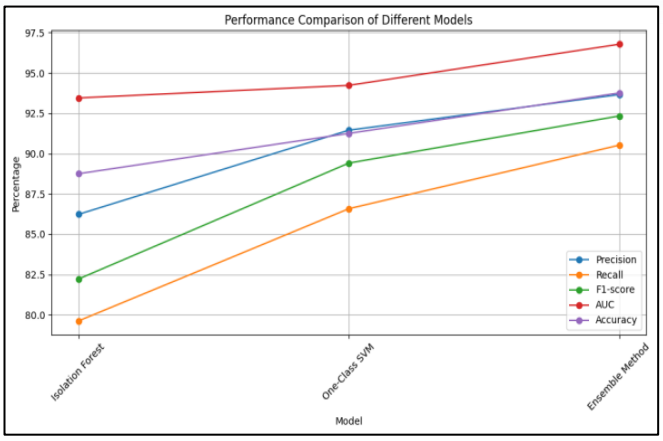

Figure 4: Performance comparison of different model

Accuracy is the percentage of properly labeled cases out of all instances. It gives a general idea of how well a model is doing, but it can be wrong when the datasets aren't fair. With 93.76% accuracy, the Ensemble Method is the best. One-Class SVM comes in second with 91.25%, and Isolation Forest comes in third with 88.75%.

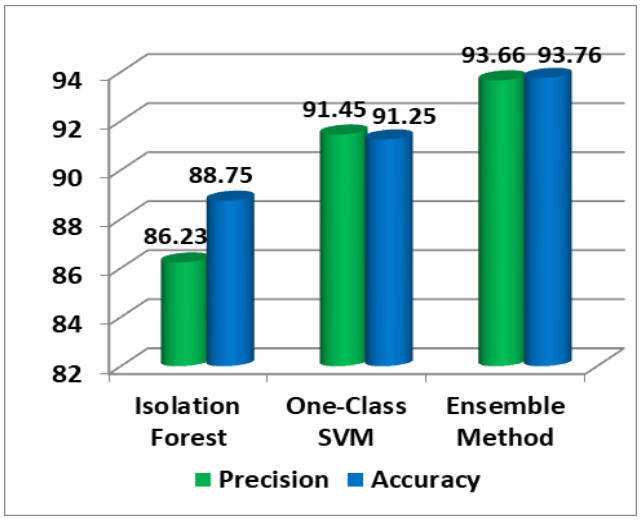

Figure 5: Accuracy and Precision comparison

The Ensemble Method does a better job than Isolation Forest and One-Class SVM in terms of precision, recall, F1-score, AUC, and accuracy. This shows that it can find integrity attacks in cyber-physical key systems.

6. Conclusion

Ensemble modeling can be used to find integrity attacks in cyber-physical key systems. This is a hopeful way to make security better and more resistant to new cyber threats. The framework is a strong way to find small changes in data that could be signs of an integrity attack. It does this by mixing the best features of several machine learning algorithms and adding anomaly detection methods. The test results show that the ensemble method is better at getting high precision, recall, F1-score, AUC, and accuracy than single models such as Isolation Forest and One-Class SVM. This great speed is very important for finding and stopping integrity threats quickly, which lowers the chance that they will stop key infrastructure activities. To make the framework better in the future, more study and development are needed to do things like finding the best ensemble model parameters, looking into new techniques for finding anomalies, and responding to new cyber dangers. Also, researchers, people in the industry, and policymakers need to work together to make sure that these advanced security solutions are used correctly in real-world cyber-physical infrastructures. This will protect them from attacks on their integrity and keep them running and reliable.

References

[1] Ghobaei-Arani, M.; Souri, A.; Baker, T.; Hussien, A. ControCity: An autonomous approach for controlling elasticity using buffer Management in Cloud Computing Environment. IEEE Access 2019, 7, 106912–106924.

[2] Shabbir, M.; Shabbir, A.; Iwendi, C.; Javed, A.R.; Rizwan, M.; Herencsar, N.; Lin, J.C.W. Enhancing security of health information using modular encryption standard in mobile cloud computing. IEEE Access 2021, 9, 8820–8834.

[3] Baker, T.; Mackay, M.; Randles, M.; Taleb-Bendiab, A. Intention-oriented programming support for runtime adaptive autonomic cloud-based applications. Comput. Electr. Eng. 2013, 39, 2400–2412.

[4] Al-Khafajiy, M.; Baker, T.; Asim, M.; Guo, Z.; Ranjan, R.; Longo, A.; Puthal, D.; Taylor, M. COMITMENT: A fog computing trust management approach. J. Parallel Distrib. Comput. 2020, 137, 1–16.

[5] Xia, T.; Washizaki, H.; Fukazawa, Y.; Kaiya, H.; Ogata, S.; Fernandez, E.B.; Kato, T.; Kanuka, H.; Okubo, T.; Yoshioka, N.; et al. CSPM: Metamodel for Handling Security and Privacy Knowledge in Cloud Service Development. Int. J. Syst. Softw. Secur. Prot. (IJSSSP) 2021, 12, 68–85.

[6] Mishra, P.; Negi, A.; Pilli, E.; Joshi, R. VMProtector: Malign Process Detection for Protecting Virtual Machines in Cloud Environment. In Third International Conference on Advances in Computing and Data Sciences, ICACDS 2019, Ghaziabad, India, April 12–13, 2019, Revised

Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages: 93 – 101 | e-ISSN: 2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.67

Selected Papers, Part I; Springer: Singapore, 2019; pp. 360–369.

[7] Ajani, S. N. ., Khobragade, P. ., Dhone, M. ., Ganguly, B. ., Shelke, N. ., & Parati, N. . (2023). Advancements in Computing: Emerging Trends in Computational Science with Next-Generation Computing. International Journal of Intelligent Systems and Applications in Engineering, 12(7s), 546–559

[8] Khorshed, M.T.; Ali, A.S.; Wasimi, S.A. A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. Future Gener. Comput. Syst. 2012, 28, 833–851.

[9] Teneyuca, D. Internet cloud security: The illusion of inclusion. Inf. Secur. Tech. Rep. 2011, 16, 102–107.

[10] King, N.J.; Raja, V. Protecting the privacy and security of sensitive customer data in the cloud. Comput. Law Secur. Rev. 2012, 28, 308–319.