Volume 9, Issue 8, August – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG1475

Support Vector Machine based Data Hacking Prediction using PMU Data

Sushma[1]
Department of Electrical and Electronics Engineering,
Ghousia College of Engineering, Ramanagara, Karnataka

Amanulla[2]
Associate Professor
Department of Electrical and Electronics Engineering,
Ghousia College of Engineering, Ramanagara, Karnataka

Javid Akthar[3]
Professor and HOD,
Department of Eletrical and Electronics Engineering,
Ghousia College of Engineering, Ramanagara, Karnataka.

Abstract:- As global reliance on power systems grows due to increasing energy demands and modern consumption patterns, maintaining the stability and reliability of the power grid has become crucial. Power systems are complex and nonlinear, and their operations are continuously evolving, making it difficult and expensive to ensure stability. Traditionally, power systems are designed to handle a single outage at a time. However, recent years have seen several significant blackouts, each originating from a single failure, which have been extensively reported. These reports are vital for mitigating operational risks by strengthening systems against identified high-risk scenarios. While extensive research has been conducted on these blackouts, cyber-attacks introduce a new dimension of risk. The advent of Phasor Measurement Units (PMUs) has enabled centralized monitoring of power system data, allowing for more effective fault and cyber-attack detection.This paper proposes a machine learning-based approach to detecting cyber-attacks using PMU data. Given the complexity and volume of power system data, traditional mathematical and statistical methods are challenging to implement. Instead, a Support Vector Classification (SVC) algorithm is used for binary classification, distinguishing between 'attack' and 'normal' states. The algorithm is trained on PMU data and evaluated using metrics such as the AUC-ROC curve and confusion matrix, achieving an 82% AUC- ROC score, demonstrating its effectiveness in identifying cyber-attacks.

Keywords:- *Cyber Attack; Support Vector Machine; AUC-ROC; Support Vector Classification.*

I. INTRODUCTION

The data transmitted from Phasor Measurement Units (PMUs) to Phasor Data Concentrators (PDCs) can be easily accessed and modified, posing significant security risks. Although previous attacks have been confined to local area networks (LANs), similar vulnerabilities can be exploited over wide area networks (WANs) such as the Internet. Research has highlighted weaknesses in the Border Gateway Protocol (BGP), which can allow attackers to reroute data packets to unintended destinations [1]. To address these risks, implementing a unique network architecture, despite its cost, is crucial. Additionally, enforcing mandatory updates for default passwords can help prevent unauthorized access.To counter these security challenges, several methods have been proposed. Principal Component Analysis (PCA) and Support Vector Machines (SVM) can be used to identify fraudulent data entries. A data- driven approach utilizing spatiotemporal relationships in PMU measurements has been suggested to differentiate between real and fake power grid events [2]. Enhancing security through bit masking has been proposed to ensure data integrity and confidentiality [3]. Developing a cybersecurity research simulation testbed within the PMU's allotted time frame has progressed. The simulation application was created by the University of Illinois at Urbana-Champaign and is both interactive and extensible. There are three customizable simulators included in this package: a PMU, a PDC, and a control center. Moreover, artificial neural networks (ANN) have been widely renowned as a highly utilized method for classification and prediction, in addition to the previously mentioned methodologies[4]. The ANN model can be represented as either a simple feed forward neural network (FNN) or a more intricate deep neural network (DNN)[5]. Their model can be obtained by solving an optimization issue, which can be efficiently tackled utilizing various local and global methods such as gradient-based search techniques [6], genetic methods [7], and others. Unsupervised learning (UL) refers to the extraction of significant patterns from unlabeled data. This process entails extracting pertinent attributes, classifications, and frameworks straight from the unprocessed data, without any manual intervention such as labeling or input

Artificial neural networks (ANNs), including both simple feedforward neural networks (FNNs) and more complex deep neural networks (DNNs), are widely used for classification and prediction. Optimization techniques such as gradient-based searches and genetic algorithms are employed to refine ANN models. Unsupervised learning (UL) methods like Isolation Forests (IF) and Autoencoders (AE) are used to detect anomalies such as false data injection

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG1475

attacks (FDIA) and Denial of Service (DoS) attacks [8], [9], [10], [11]. Dynamic Bayesian Networks (DBN) are also utilized for attack detection [12]. Semi- supervised learning (SSL) combines labeled and unlabeled data to enhance detection capabilities. Techniques like semi- supervised adversarial autoencoders (SSAA) and generative-adversarial frameworks are proposed for improved FDIA detection, with new models such as SS-deep-ID and robust semi-supervised prototypical networks (RSSPN) offering advanced detection methods (References [13], [14], [15], [16].

II. METHODOLOGY

➢ *PMU Dataset*

The dataset employed for classification consists of various features, as detailed in Table 1. It encompasses 128 attributes, with the target variable denoting whether the measurement pertains to a 'fault' or a 'normal' event. The data originates from Phasor Measurement Units (PMUs), which are sophisticated devices designed to capture and compute electrical waveforms on the power grid by synchronizing with a standard time reference.Each PMU records 29 distinct types of measurements, resulting in a total of 116 measurement columns across four PMUs. Additionally, the dataset includes three types of logs: relay logs, control panel logs, and Snort logs. Relay logs document the activities of protective relays that monitor electrical parameters and initiate protective measures to ensure system safety. Control panel logs capture activities and statuses from control panels that oversee and manage the power system. Snort logs come from an open-source network intrusion detection system that tracks and analyzes network traffic for malicious activities.In total, the dataset includes 128 attributes: 116 from the PMU measurements and 12 from the logs. This comprehensive dataset is used to train models to classify whether events are "normal" or indicative of an "attack," with the classification target labeled as "Marker."

➢ *SVC based Detection Algorithm*

The attack events are due to different cyber attack that can happen to a power system which include the data injection in the power system that may cause the relay to operate without actually having any fault in the power system. Remote tripping fault is the one which would trip the relay without any event occurring just by the cyber-attack. This is called the command injection attack type. Two subtypes of this attack is command injection in single relay and in multiple relays. Then the third type of the cyber attack is the relay setting change attack. This is by making the relay not to act even when the fault is available. These 128 features or variables primarily originate from synchrophasors or phasor measurement units (PMUs). The data was sampled at 120 samples per second, with each scheme simulated for 17 seconds. Different fault details are shown in Table 1. The fault prediction process comprises four key components: Data Preprocessing Automation, Outlier Detection and Feature Engineering, Training and Testing, and Model Evaluation. The Support Vector Machine (SVM) method is utilized to enhance the model's generalization capabilities. Data preprocessing involves several automated steps, including anomaly detection, data cleaning, and the organization of data into balanced and unbalanced datasets. This process establishes the framework for the fault prediction model. Automated procedures address data impurity and missing values, with mean values used to replace missing entries. Given the critical role of fault prediction in electrical systems, ensuring the reliability of the prediction algorithm is paramount.

To handle large volumes of data effectively, the method must offer strong generalization and utilize highly orthogonal inputs.

Advanced feature engineering techniques may be needed to improve prediction accuracy, especially if the data exhibits significant correlations. Developing a data-aware preprocessing strategy is complex but essential.The workflow includes dividing the dataset into training and testing subsets, each containing relevant CSV files. For the PMU cyberattack detection, the target variable indicates whether an attack has occurred. The machine learning model's objective is to predict if the PMU data suggests a cyberattack.

Table 1 Attack Event Scenarios in Power System

Scenario	Attack Type
	Data Injection
	Attack Sub-type (SLG fault replay)
7	Fault from 10-19% on L1 with tripping command
8	Fault from 20-79% on L1 with tripping command
9	Fault from 80-90% on L1 with tripping command
10	Fault from 10-19% on L2 with tripping command
11	Fault from 20-79% on L2 with tripping command
12	Fault from 80-90% on L2 with tripping command
	Remote Tripping Command Injection
	Attack Sub-type (Command injection against single relay)
15	Command Injection to R1
16	Command Injection to R2
17	Command Injection to R3
18	Command Injection to R4
	Attack Sub-type (Command injection against single relay)
19	Command Injection to R1 and R2
20	Command Injection to R3 and R4
	Relay Setting Change
	Attack Sub-type (Disabling relay function - single relay disabled & fault)
21	Fault from 10-19% on L1 with R1 disabled & fault
22	Fault from 20-90% on L1 with R1 disabled & fault
23	Fault from 10-49% on L1 with R2 disabled & fault
24	Fault from 50-79% on L1 with R2 disabled & fault
25	Fault from 80-90% on L1 with R2 disabled & fault
26	Fault from 10-19% on L2 with R3 disabled & fault
27	Fault from 20-49% on L2 with R3 disabled & fault
28	Fault from 50-90% on L2 with R3 disabled & fault
29	Fault from 10-79% on L2 with R4 disabled & fault
30	Fault from 80-90% on L2 with R4 disabled & fault

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG1475

III. RESULTS AND DISCUSSION

To generate box plots for the first 14 columns of numerical data from a dataset containing Phasor Measurement Unit (PMU) data the sea born library from python is used and they are shown as follows. Since it is a classification algorithm the amount of majority and minority class has to be checked whether it is balanced or imbalanced . The class distribution graph for the PMU considered is as given in the Figure 1

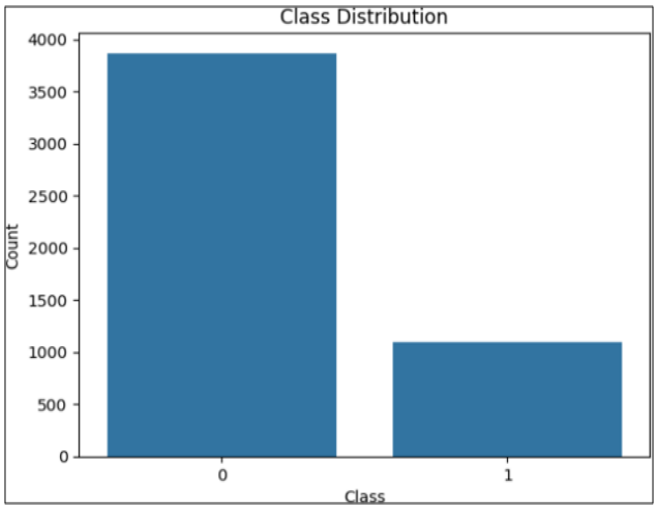

Fig 1 Class Distribution Graph

The confusion matrix is a fundamental tool for assessing the performance of classification models. It provides detailed insights into how well the model predicts each class, helping identify strengths and weaknesses. By understanding and utilizing the confusion matrix, you can make informed decisions about model improvements and better evaluate its effectiveness. The confusion matrix for the implementation of cyber attack detection is as given in the Figure 2. The label '1' is the Natural class and '0' is the attack class. Since the attack class is the minority class the samples of the majority class was brought to be equal to the majority class to make it a balanced data.

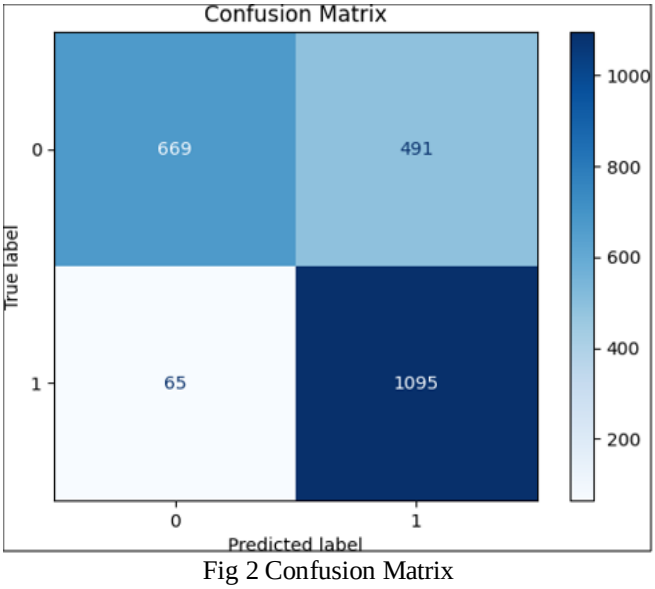

Fig 2 Confusion Matrix

The performance of the cyber attack detection implementation is found to be satisfactory with 0.82 as the area under the curve. It is a measure of how many correct classification can happen in the machine learning algorithm. It infers that above 80% of the classification is correct. On further tuning the algorithm the performance can be improved. From the analysis thus developed the performance metrics is as given in the table 2.

Table 2 Performance Metrics

Accuracy	0.76
Precision	0.69
Recall (Sensitivity):	0.94
F1 Score	0.80
Specificity	0.58

IV. CONCLUSION

This work presents a machine learning-driven approach to cyber-attack detection in power system. Support vector classifier based implementation is carried out to classify the events from the PMU data gathered from the power system. The dataset of the PMU data having 32 attributes from four such PMUs are used in the prediction implementation. The imbalance in the data is treated by taking the majority class data to be equal to the number of minority class data for better performance. The findings demonstrate that the support vector machine approach greatly enhances performance in the identifying cyber attack detection. The method's remarkable 82% accuracy rating underscores its promise for dependable detection algorithm.

REFERENCES

[1]. internet.http://www.wired.com/threatlevel/2013/12/bgp-hijacking-belarus-iceland/. Accessed: 2022-12-15.
[2]. Q. Sun, L. Shi, Y. Ni, D. Si, and J. Zhu, ''An enhanced cascading failure model integrating data mining technique,'' Protection Control Mod. Power Syst., vol. 2, no. 1, pp. 209–219, Jan. 2017.
[3]. R. Vijayanand, D. Devaraj, B. Kannapiran, and K. Kartheeban, ''Bit masking based secure data aggregation technique for Advanced Metering Infrastructure in Smart Grid system,'' in Proc. Int. Conf. Comput. Commun. Inform., Jan. 2016, pp. 45–54.
[4]. T. J. Overbye, Z. Mao, K. S. Shetye, and J. D. Weber, ''An interactive, extensible environment for power system simulation on the PMU time frame with a cyber security application,'' in Proc. IEEE Power Energy Conf., Feb. 2017, pp. 1–6.
[5]. Z. Mao, T. Xu, and T. J. Overbye, ''Real-time detection of malicious PMU data,'' in Proc. Int. Conf. Intell. Syst. Appl. Power Syst., Sep. 2017, pp. 121–128.
[6]. https://www.kaggle.com/bachirbarika/power-system

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG1475

[7]. S. Wang, M. Roger, J. Sarrazin et al., "Hyperparameter optimization of two-hidden-layer neural networks for power amplifiers behavioral modeling using genetic algorithms," IEEE Microwave and Wireless Components Letters, vol. 29, no. 12, pp. 802-805, Dec. 2019
[8]. S. Ahmed, Y. Lee, S. Hyun et al., "Unsupervised machine learningbased detection of covert data integrity assault in smart grid networks utilizing isolation forest," IEEE Transactions on Information Forensics and Security, vol. 14, no. 10, pp. 2765-2777, Oct. 2019.
[9]. J. Wang, D. Shi, Y. Li et al., "Distributed framework for detecting PMU data manipulation attacks with deep autoencoders," IEEE Trans- actions on Smart Grid, vol. 10, no. 4, pp. 4401-4410, Jul. 2019.
[10]. M. Aboelwafa, K. Seddik, M. Eldefrawy et al., "A machine-learningbased technique for false data injection attacks detection in industrial IoT," IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8462-8471, Sept. 2020.
[11]. K. Lu, G. Zeng, X. Luo et al., "Evolutionary deep belief network for cyber-attack detection in industrial automation and control system," IEEE Transactions on Industrial Informatics, vol. 17, no. 11, pp. 7618-7627,Nov. 2021
[12]. I. Sohn, "Deep belief network based intrusion detection techniques: a survey," Expert Systems with Applications, vol. 167, pp. 1-9, Apr. 2021.
[13]. Y. Zhang, J. Wang and B. Chen, "Detecting false data injection at- tacks in smart grids: A semi-supervised deep learning approach," IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 623-634, Jan. 2021.
[14]. M. Farajzadeh-Zanjani, E. Hallaji, R. Razavi-Far et al., "Adversarial semi-supervised learning for diagnosing faults and attacks in power grids," IEEE Transactions on Smart Grid, vol. 12, no. 4, pp. 3468-3478, Jul. 2021.
[15]. M. Abdel-Basset, H. Hawash, R. Chakrabortty et al., "Semi-super- vised spatiotemporal deep learning for intrusions detection in IoT net- works," IEEE Internet of Things Journal, vol. 8, no. 15, pp. 12251-12265, Aug. 2021
[16]. T. Zheng, Y. Liu, Y. Yan et al., "RSSPN: robust semi-supervised proto- typical network for fault root cause classification in power distribution systems," IEEE Transactions on Power Delivery, Nov. 2021. DOI: