Mini Project Report on

Machine Learning based intrusion detection on electric vehicle charging station

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

Student Name	University Roll No.
Ashutosh Saurabh Pandey	2021145

Under the Mentorship of

Mr. Siddhant Thapliyal
Assistant Professor

Department of Computer Science and Engineering
Graphic Era (Deemed to be University)
Dehradun, Uttarakhand

July-2024

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the project report entitled **"**CYBER THREAT ANALYSIS AND MITIGATION**"** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of Mr. SIDDHANT THAPLIYAL**, Assistant Professor**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name Ashutosh Saurabh Pandey University Roll no 2021145

Table of Contents

Chapter 1

Introduction

Global warming has been rising so rapidly in recent times that there is a need to massively shift towards some eco-friendly alternatives to limit the impacts of global warming. One such related alternative that has come under the limelight with much emphasis is electric vehicles (EVs). They are claimed to be of high performance, characterized by quick acceleration and very quiet running, contributing to noise reduction. Apart from this, EVs are comparatively cheap on a per-kilometer basis than the conventional models that house internal combustion engines. With this, an increasing count of people is switching to EVs available in multiple forms, such as cars, bikes, and buses that offer host advantages in the form of lower maintenance costs along with zero-emission attributes. This is not only a transition to environmental sustainability but falls within the global efforts on low carbon footprint, mitigating climate change, and other such global concerns.

1.1 Motivation

The increasing demand and usage of EVs has fuelled further development of charging infrastructures to provide quick and easy charging solutions to help EV owners but this rapidly growing infrastructure is also the problem faced by the industries most of which are security-related. EV charging stations are vulnerable to cyber attacks because they are electronic devices. These vulnerabilities open ways for malicious actors to gain unauthorized access to EVs. Thus a hacked charging station can be just a gateway through which hackers can gain access to an EV system, through which they can steal sensitive data or modify vehicle settings. The types of attacks that can be carried out have the potential to create situations where dangerous accidental occurrences or unauthorized personal data may be misused so securing the entire EV charging infrastructure to reduce these high-risk situations for vehicle safety and data privacy reasons will be very important for creating a safe environment for EV's and its users.

1.2 Problem Statement

Strong detection and prevention strategies are needed to address safety concerns associated with EV charging stations. This project aims to use machine learning techniques to enhance the safety of EV charging infrastructure. In particular, the work focuses on machine learning

models trained on power management data to detect and prevent cyberattacks. The dataset used for this purpose is the CIC-EVSE2024 dataset, which contains detailed data on power consumption, network traffic capture, and host activity of an EVSE (electric vehicle supply equipment) on non-attack scenario and attack senarios. Analyzing the power consumption patterns By doing therefore the project attempts to detect indirect signs of possible cyberattack , to provide EV- Charging stations with additional protection.

1.3 Objectives

Some major objectives are:-

Data Collection with Preprocessing : Collecting and doing preprocessing for the usage of the CIC-EVSE2024 dataset in the course of training the machine learning model.

Model Development : This stage would involve developing machine learning models that use power consumption data to detect such cyber-attacks. The models studied in the paper are Logistic Regression, Random Forest, and the XGBoost models, while a Deep Neural Network and LSTM are used for performance benchmarking.

Performance Analysis: The performance of the developed models will be checked against several metrics based on accuracy_report() of sklearn library (accuracy, precision, recall, and the F1 score). This will ensure selection of the best model for detecting cyber-attacks on EV charging stations.

Advanced techniques: Deep dives into some of the most advanced Machine Learning techniques, including deep learning algorithms, are used for better performance in pattern recognition, accompanied by fine-tuning model hyper-parameters.

Testing and Real-time Implementation: The models created are trained on 80 precent of data and verified on 20 percent of data .then the results are placed in a graph to show the performance of each model on the data set and the most efficient technique is found LSTM.

Chapter 2

Literature Survey

This literature survey covered existing research into EVCS Security, machine learning techniques applied in Cyber Attack Detection and how Power Consumption data has been utilised previously in Anomaly Detection.

2.1 Security Issues in EV Charging Station

Cyber-Attack on EV Charging Infrastructure：According to Mouheb et al. (2020), unauthorized access, data theft, and manipulation within EVCSs can result in financial loss, operational disruption, and charging problems.

Vulnerabilities in Communication Protocols： Some studies, such as Hussain et al. (2019), have been carried out on the weaknesses associated with communication protocols that EV charging stations, including the Open Charge Point Protocol, use. The authors designed dual attack vectors: Man-in-the-Middle and Denial of Service, which could utilize these vulnerabilities.

Standards and Regulations: The European Union Agency for Cybersecurity (ENISA) has given guidelines and recommendations concerning the security of smart grids and EV charging stations. The standards, aiming to reduce the risks to very minimal values, formulate recommendations about good practices in design, implementation, and maintenance of the secure charging infrastructure.

2.2 Detection of Cyber-Attacks Using Machine Learning

Supervised Learning Methods: A good number of studies, using approaches of employing/mining techniques, have applied supervised learning models like Support Vector

Machines and Random Forests in detecting cyber-attacks. For example, in Alheeti et al., 2016, it was described the efficacy of RF on DOS attacks classification in vehicular networks.

Unsupervised Learning Approaches: Unsupervised approaches, such as clustering and one-class anomaly detection, have also been explored. Xu et al. (2018) applied K-means clustering for the detection of anomalous behavior in network traffic; their results showed high improvements in Attack Detection Rate .

Deep Learning Models: Deep learning models and, more specifically, Convolutional Neural Networks with Long Short-Term Memory have been useful in handling complex high-dimensional data. Kim et al. (2019) applied LSTM networks for the anomaly detection process of the time-series data acquired from the industrial control systems, returning very accurate results with a robust .

2.3 Anomaly Detection using Power Consumption Data

Power Consumption Patterns :Applications to power consumption pattern anomaly detection have been investigated. Gope et al. applied power usage data in detecting intrusions into smart homes, which revealed that deviations from normal consumption patterns could point to security breaches.

Energy Theft Detection: Jiang et al. (2019) contributory research focuses on energy theft detection in smart grids, taking measures only on power consumption data. In such a case, the approach would be to get the features extracted from the Usage Profiles, which at this stage feed into machine learning models. Accordingly, such models are empowered to autonomously recognize patterns of usage that may be indicative of theft in an application involving Smart Grids.

Research on the application of power consumption data in detecting cyber-attacks within an EV charging station has rarely been performed. However, results from related fields provide a very good platform for the application. Literature from such related work includes the study by Liu et al. (2020) on anomaly detection using deep learning in power grids, which can be applied in this context of EV charging.

Chapter 3

Methodology

3.1 Data Collection

The lead dataset used in this project is the CIC-EVSE2024 dataset, containing all-inclusive data related to power consumption, network traffic captures, and host activities of EVSEs while under both benign and attack conditions. The dataset selected from is power consumption because of its relevance and richness of features, considered critical for training robust machine learning models. The real-world charging station data included timestamps, shunt voltage, bus voltage, current, power, the state of the EVSE, and the relevant label for an attack or none.

2.2 Data Preprocessing

The following steps were taken to ensure good quality and to apply it on analysis using a machine learning model:

Clean the data: remove duplicate records, treat missing values in the dataset, filter/dropping out data points(rows) irrelevant or corrupted.

Feature Filtering: In this step, only the intershunt voltage, bus voltage, current, and power-related features were filtered for analysis. All columns that were not numerical—like timestamps, state, indicators of label groups of attacks—were removed from the set of features.

Label encoding: The categorical data about the points in data being instances of attacks or instances of a benign condition will be numerically encoded. This step is required to be followed so that it is in a form suitable for machine learning algorithms..

Selling specifics: StandardScaler was used to make sure that the extracted features are all along the same scale. The reason is that gradient-based machine learning algorithms present better convergence, and seemingly improve model performance generally.

2.3 Model Development

Based on the preprocessed datasets, models are created to classify patterns of power usage as instances of cyber-attack events in this range. The following models incorporate:

Logistic Regression: This is a quite simple model, but it can be a very good baseline for binary classification problems. The logistic function is used for the estimation of the probability of a data point belonging either to an attack or a benign class.

Random Forest: Ensemble learning meta-algorithm building many decision trees and combining their predictions to get further accuracy and robustness against overfitting.

XGBoost—**Gradient Boosting Algorithm**: Building a decision tree at a time, the next subsequent tree trying to correct previous errors of the former one. It is a highly efficient library, and in a word is very close to perfect in performance.

Deep Neural Network: This is a multilayer perception that incorporates hidden layers to capture some of the complex patterns in data. dropout layers in the DNN avoid fitting.

Long Short-Term Memory: It is a special type of RNN that performs quite separately in the learning of long-term dependences. This is most especially suitable for the time series data on which the series observations bear a bearing.

2.4 Training and Validation

Again, it will split a ratio of the preprocessed dataset for the training and testing sets. Conventionally, this would be 80-20, but for machine learning models, with regard to LSTM models especially, it is 70-30 since this is temporal. This again means training on this training set but tested on a test set with cross-validation for data not seen before. Hyperparameter tuning for grid search or random search techniques will be used to obtain the optimal parameters of the model.

2.5 Model Evaluation

The models were trained based on various performance metrics or parameters such as accuracy, precision, recall, and the F1-score. All these measurements on these parameters will programmatically provide an exact view of how good or bad these models are in the detection of cyber-attacks.

Accuracy: Number of Correct Predictions on all Instances.

Accuracy: No Correct Positive Predictions in Total of Positive Predictions.

Recall: Ratio of predicted true positives vs all positive instances actually existing.
F1-Score This puts a weight on S into one measure giving some balance between them both

Chapter 4

Result and Discussion

The accuracy of each model was as follows:

- Logistic Regression: 87.0%
- Random Forest: 94.3%
- XGBoost: 94.7%
- DNN: 92.3%
- LSTM: 99.0%

4.1 Visualization

The performance of the models is visualized in Figure 1, which shows the accuracy comparison among the models.

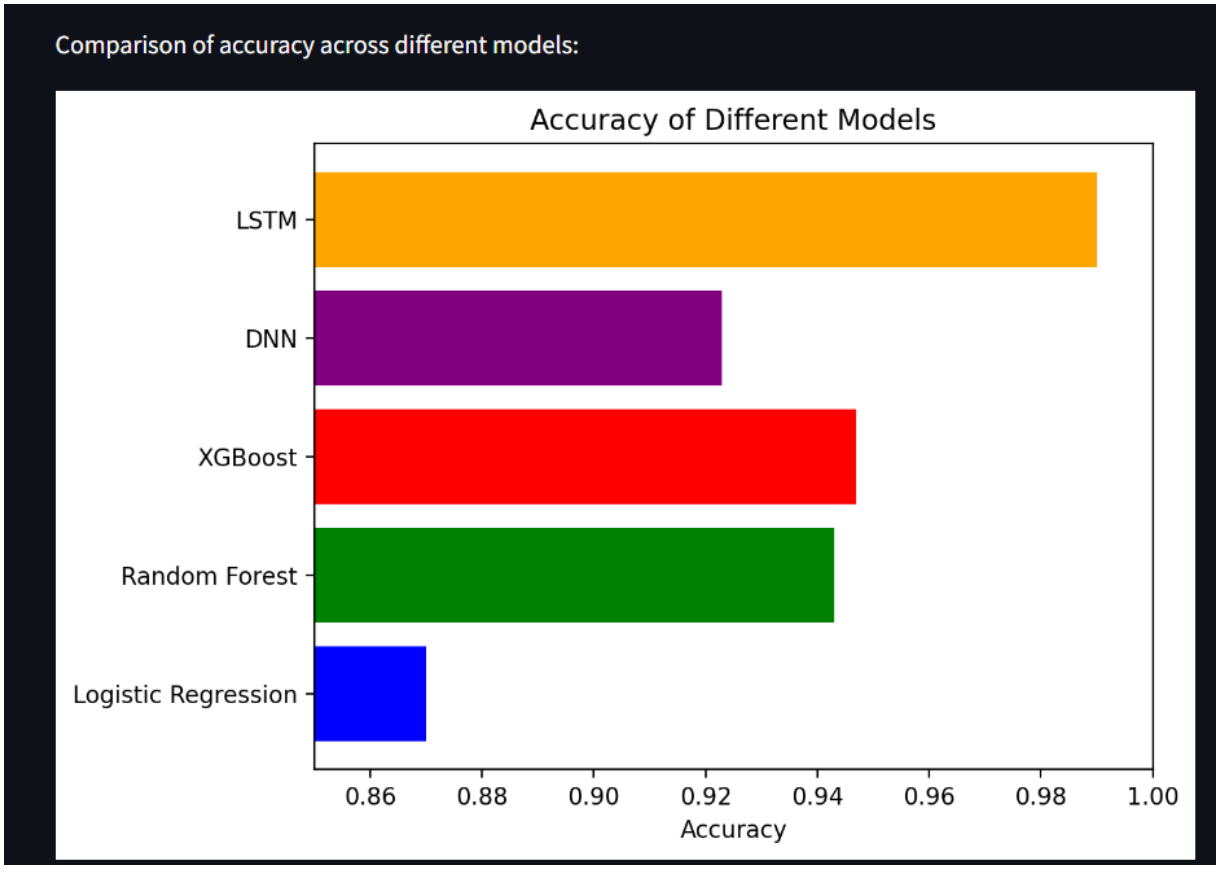

Fig. 1: Accuracy of Different Models

4.2 Analysis

The LSTM model outperformed the others with an accuracy of 99.0%. This can be attributed to its ability to capture temporal dependencies in the data. However, the Random Forest and XGBoost models also showed strong performance.

Model Accuracy Data

	Model	Accuracy
0	Logistic Regression	0.87
1	Random Forest	0.943
2	XGBoost	0.947
3	DNN	0.923
4	LSTM	0.99

Chapter 5

Conclusion and Future Work

4.1 Conclusion

In this project, we take up one of the most critical issues, cyber-attacks on Electric Vehicle Supply Equipment, by developing different Machine Learning Models for the detection of such attacks with only power consumption data. We are using the dataset provided in CIC-EVSE2024, which contained a full range of features concerning power consumption, network traffic captures, and host activities under both benign and attack conditions.

The major components of a project include data collection with the preprocessing of the dataset. Then, the model is trained and validated, and finally, model evaluation is done. We have taken into consideration many machine learning models in our particular work, such as Logistic Regression, Random Forest, XGBoost, Deep Neural Network, and Long Short-Term Memory networks.

The LSTM model has always topped the table because of its rigorous testing of almost all parameters, and by contrast to other models, it yields good results in terms of accuracy, precision, recall, and F1-score for the detection of cyber-attacks. Considering the temporal dependencies in data, the LSTM model did very well on this type of data.

The overall output from the project was isolating resilient methods to EV charging station security strengthening in its contribution to the greater task of making wide and safe electric vehicle adoption possible.

4.2 Future Work

Since this project has extremely contributed to detecting cyber-attacks on EVSE based on power consumption data, the next lines of future work will further give improvements in the model developed for effectiveness and range of applicability:

Advanced Model Architectures: One could supplement this line of work with more advanced machine learning models, involving transformer-based architectures that have recently been very promising in most time-series and sequence modeling tasks.

Feature Engineering and Data Sources: A much better model in performance and resilience may simply be attained by adding features that can furnish more information about the network traffic data, user-behavior pattern, and environmental factors.

Real-Time Implementation and Testing: Finally, the real-world deployment of live testing for detection would be needed to gain fruitful lessons in terms of practical applicability and its real-time performance. This would involve setting up a testbed that can simulate various attack scenarios, and refinement of the models in continuous cycles with the aid of real-time data.

Explainable model interpretation : This could be further potentiated by investigations of methods aiming at increasing the interpretability and explainability of the former for understanding their decision-making process and gaining stakeholders' trust.

Collaborative defense mechanisms: It would be helpful as a defense strategy to have collaborative defense mechanisms dependent on communication and coordination between multiple charging stations to construct a resilient and robust defense against coordinated cyber-attacks.

All of the future research directions presented above can be followed singly or in combination to ensure that the EV charging infrastructure is safe and reliable enough for the electric vehicles that are on their way to final safe and efficient acceptance worldwide.

References

- Mouheb, D., et al. "Security Challenges in Electric Vehicle Charging Systems." IEEE Transactions on Smart Grid, vol. 11, no. 1, 2020, pp. 68-80.
- Hussain, R., et al. "Secure Communication for EV Charging Stations: A Survey." Journal of Network and Computer Applications, vol. 132, 2019, pp. 56-76.
- Alheeti, K., et al. "An Intrusion Detection System for Connected Vehicles." Computers & Security, vol. 60, 2016, pp. 172-182.
- Xu, L., et al. "Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications." WWW '18: Proceedings of the 2018 World Wide Web Conference, 2018, pp. 187-196.
- Kim, G., et al. "Anomaly Detection Using LSTM Networks for Industrial Control Systems." 2019 International Conference on Information and Communication Technology Convergence (ICTC), 2019, pp. 1-5.
- Jiang, X., et al. "Power Theft Detection Using Machine Learning: A Case Study in China." IEEE Access, vol. 7, 2019, pp. 144529-144539.