Machine Learning Based Intrusion for Internet of Things(IoT)

Mohammed Aman(2020A7PS0050U)

Supervisor: Raja Muthalagu

ABSTRACT

Machine Learning models used include Decision Tree Classifier, Random Forest Classifier & Logistic Regression. The train and test split is done such that 80% of the dataset consisting of 7mil+ entries goes to training and the rest 20% is allocated for testing. The dataset contains Network and intrusion details of common IoT devices. The performance metrics used are Accuracy(Ac), Recall(Rc), Precision(Pr), F1 score(F1s).

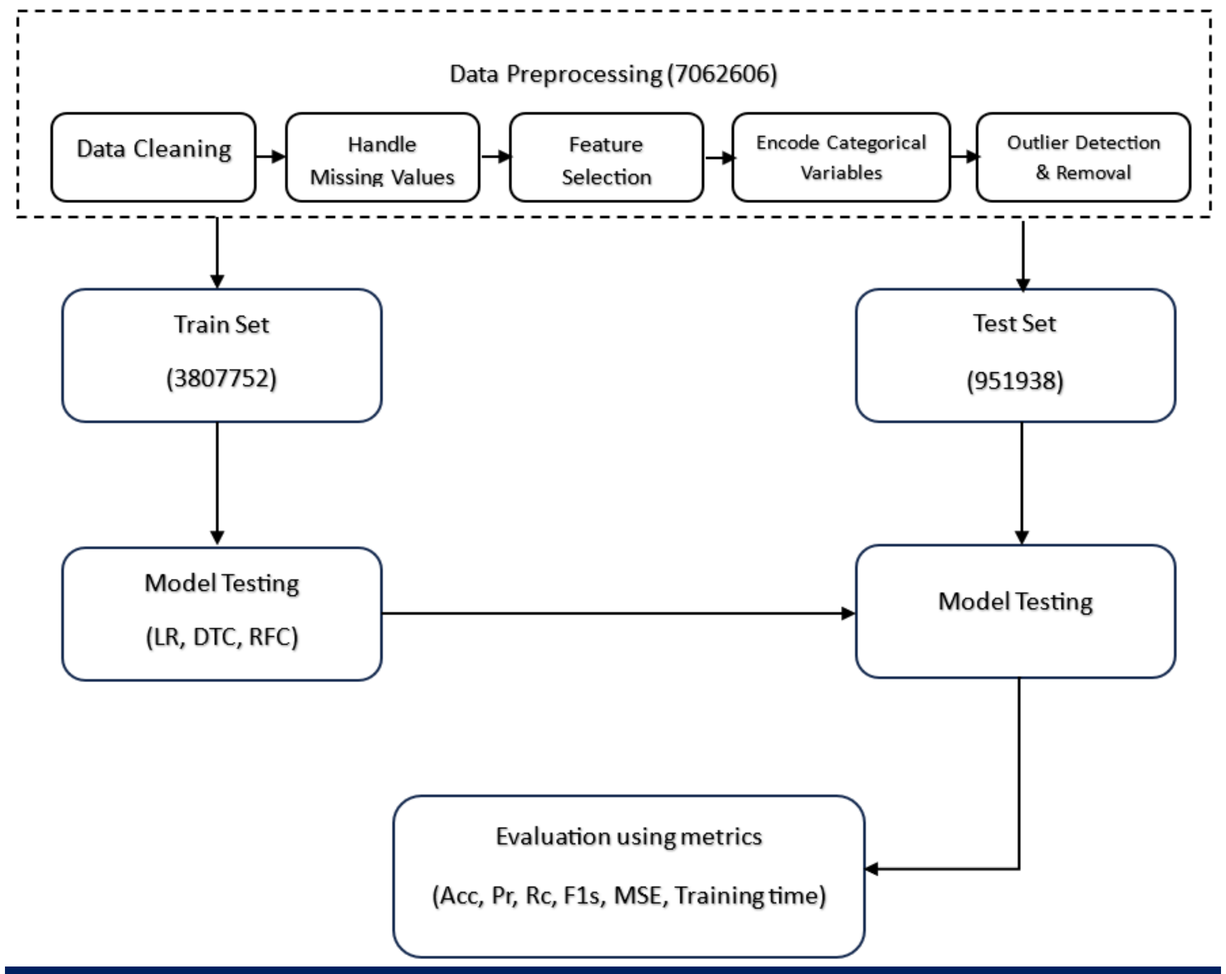
OBJECTIVES

- To understand Network Intrusion and its strategies
- Implement a variety of machine learning models, including Logistic Regression(LR), Decision Tree Classifier(DTC), and Random Forest Classifier(RFC), to detect network intrusion and to compare these techniques for finding ideal model for implementation.

INTRODUCTION

- The rise of Internet has caused an increase in the amount of information traffic which has caused security concerns. IDS detects such attacks on computers and networks. This project aims to establish IDS on IoT devices.
- Many modern IDS are ML models as they can handle a large volume of network and data. However, this project aims to find the best ML model for the same by comparing different models to a dataset consisting of 7mil+ entries of IoT Devices.

IMPLEMENTATION



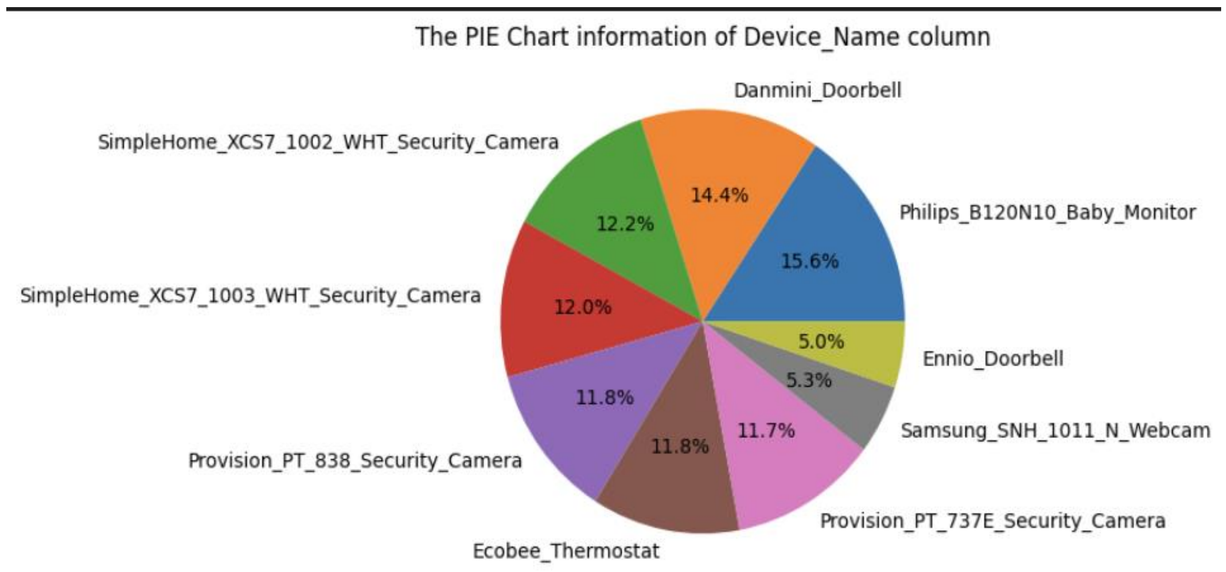
DATASET

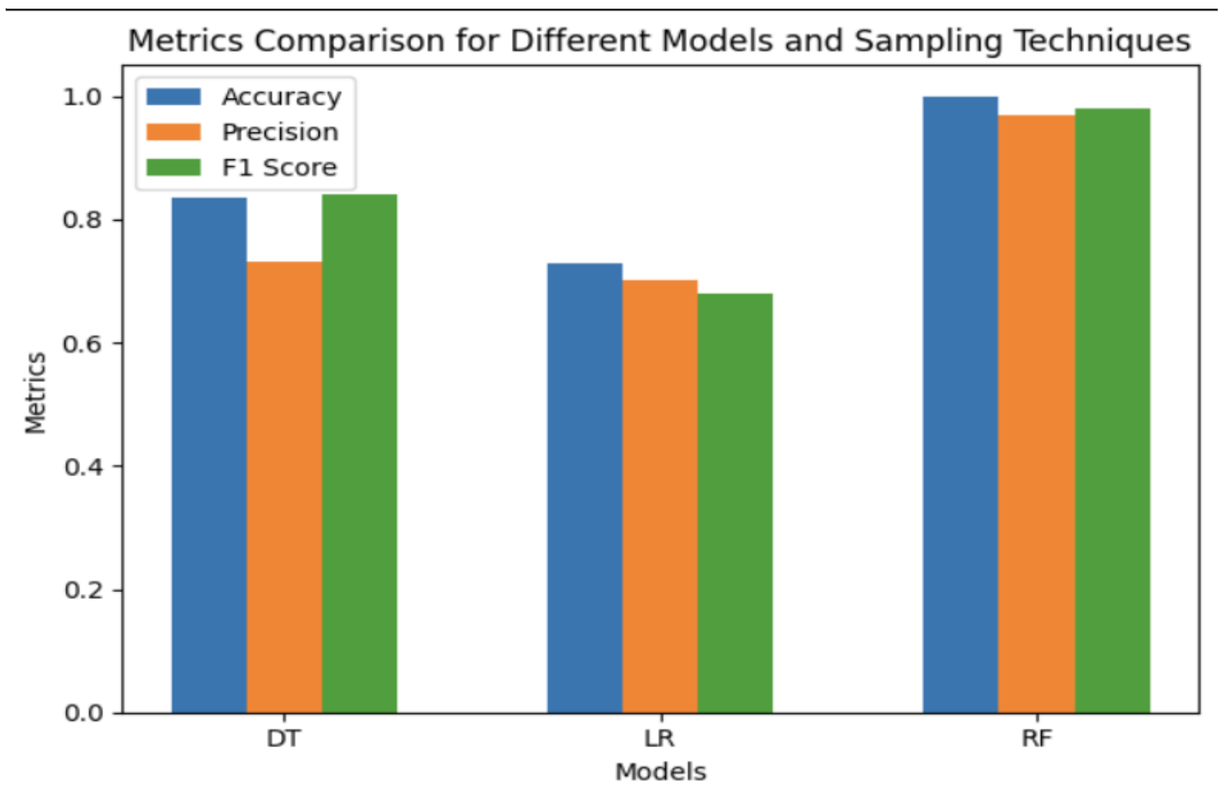
The BoTNeTIoT-L01 dataset was used containing nine IoT devices traffic sniffed using Wireshark in a local network using a central switch. Initially consisting of 7062606 entries, it was narrowed down to 4759690 entries out of which 3807752 was used for training and 951938 was used for testing.

RESULTS

The classification models, including Logistic Regression, RFC, and Decision Tree, were trained and evaluated. They were assessed using metrics such as accuracy, precision, and the confusion matrix.

Additionally, a Random Forest Classifier was trained and evaluated, achieving high accuracy on the test set. Cross-validation scores further validated the model's robustness. Subsequently, a bar plot was generated to compare the performance metrics (Accuracy, Precision, and F1 Score) of different models.





Reference	Outperforming Model	Accuracy (%)
Ref. [1]	XGBoost	97%
Ref. [2]	Random Forest	99.971%
Ref. [3]	Random Forest	99.986%
Ref. [4]	KNN	92.29%
Proposed Model	**Random Forest**	**99.99968%**

CONCLUSION AND FUTURE WORK

- The analysis of the BoTNeTIoT-L01-v2 dataset revealed promising results with RandomForestClassifier outperforming other models. The model demonstrated high accuracy, precision, and F1 Score. Future work could explore fine-tuning hyperparameters and incorporating advanced techniques like ensemble methods for further improvement. Additionally, addressing class imbalance and optimizing feature selection methods could enhance model performance.

REFERENCES

[1] Z. Liu, N. Thapa, A. Shaver, K. Roy, X. Yuan and S. Khorsandroo, "Anomaly Detection on IoT Network Intrusion Using Machine Learning,"

[2] M. M. Alani, "IoTProtect: A Machine-Learning Based IoT Intrusion Detection System,"

[3] Sripad Karthik, 2023, September. “IOT data for Intrusion detection using ML and DL.”

[4] S. S. Swarna Sugi and S. R. Ratna, "Investigation of Machine Learning Techniques in Intrusion Detection System for IoT Network,"