



Article Title: IoT Network Traffic Classification Using Machine Learning Algorithms: An Experimental Analysis

IoT Network Traffic Classification Using Machine Learning Algorithms: An Experimental Analysis

Evelyn Lisa. E¹, T. L. Ajeesha M. E²

¹ Student, ² Associate Professor

Department of Applied Electronics, Loyola Institute of Technology and Science,

¹evelynsujith@gmail.com, ajeesha@gmail.com

ABSTRACT

In this project, classification of IoT network traffic using random forest classifier is proposed. More-dataset is utilized to separate the IoT traffic classification. Data preprocessing is the concept of changing the raw data into a clean data set. The dataset is preprocessed in order to check missing values, noisy data, and other inconsistencies before executing it to the algorithm. Extraction takes place using Chi-square based feature extraction. To improve classification, the extraction regions are processed using the Chi-Square technique to extract several features and select the necessary features. This project uses the effective Chi-square technique to find feature extraction and feature selection. Finally, the selected features are fed into an extra tree and random forest classifier for accurate classification. This project is implemented using Python software.

1 Introduction

The Internet of Things (IoT) is a network of physical objects equipped with sensors, software, and other technologies to communicate and share data with other devices and systems over the Internet. IoT is a phenomenon that has taken off in recent years, with the number of smart devices being forecasted at around 30.73 billion, which would mark a growth of 15 percent (Statista, 2020). This would encompass various devices such as voice controllers, doorbell cameras, smart TV, security cameras and so on, which add an extra dimension to everyday life. However, as the number of IoT devices grows, the number of security vulnerabilities from the edge to the cloud increases. IoT generally refers to an extensive network, which is practically impossible to visualize from a standalone point of view. This makes the system harder to monitor while leaving loopholes in the security protocols that can be easily bypassed to allow unfettered access to vast amounts of confidential data. The system provides little control with no evident security measures. With the rising number of IoT based tools, it is noticeable that the devices are prone to various sorts of attacks. This is due to unique constraints like multiple technologies, multiple verticals, and resource limitations that include low memory and low computational power.

2 Related Works

To perform the efficient identification of traffic using neural network classifier. To extract



Article Title: IoT Network Traffic Classification Using Machine Learning Algorithms: An Experimental Analysis

the features with the help of Chi-square method. The goal of random forests is to transform a set of high-variance, low-bias decision trees into a model with low variance and low bias.

2.1 Research Methodology

The methodology adopted in this paper comprises of four stages: Data Pre- Processing, Feature Selection Chi-Square, Extra Tree Classifier, and Random Forest Classifier

2.2 Data Pre-Processing

Data preprocessing, a subset of data preparation, refers to any type of processing performed on raw data in order to prepare it for further processing. It has traditionally been a critical first step in the data mining process. The need for data preprocessing exists because good data is unquestionably more important than good models, and the quality of the data is critical. Therefore, companies and individuals invest a lot of their time in cleaning and preparing the data for modeling. The data present in the real world contains a lot of quality issues, noise, inaccurate, and not complete.

2.3 Feature Selection Chi-Square

Feature selection, also known as attribute selection, is the process of extracting the most relevant features from a dataset and then using machine learning algorithms to improve the model's performance. A large number of irrelevant features exponentially increases training time and increases the risk of over fitting. For categorical features in a dataset, the Chi-square test is used. Calculate the Chi-square between each feature and the target and choose the features with the highest Chi-square scores. It determines whether the sample's association between two categorical variables reflects their true association in the population.

2.4 Extra Tree Classifier

Extra trees (short for extremely randomized trees) is an ensemble supervised machine learning method that uses decision trees and is used by the Train Using AutoML tool. Decision trees classification and regression algorithm for information about how decision trees work. The extra trees algorithm, like the random forests algorithm, creates many decision trees, but the sampling for each tree is random, without replacement. This creates a dataset for each tree with unique samples. A specific number of features, from the total set of features, are also selected randomly for each tree. The most important and unique characteristic of extra trees is the random selection of a splitting value for a feature.

2.5 Random Forest Classifier

Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification



Article Title: IoT Network Traffic Classification Using Machine Learning Algorithms: An Experimental Analysis

and regression problems. The random forest algorithm is an extension of the bagging method as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees. The random forest classifier can be used to solve for regression or classification problems.

3 Existing System

This existing system presents a machine learning based classification of IoT network traffic. Data preprocessing is the process of digitizing data from the world around us so it can be displayed, analyzed, and stored in a computer. Data preprocessing transforms the data into a format that is more easily and effectively processed in machine learning. The Chi-square Method is used to carry out the feature extraction process. Effective classification is accomplished with the KNN and MLP classifier, and predicted results are accurate. This approach achieves good accuracy but takes large training times in packet level due to large amounts of data and unbalanced data.

4 Proposed System

This project proposes a classification of IoT network traffic using random forest classifier. The process of digitizing data from the world around us so that it may be displayed, analysed, and stored in a computer is known as data preprocessing. Data preprocessing transforms the data into a format that is more easily and effectively processed in machine learning. The Chi-square Method is used to carry out the feature extraction process. Effective classification is accomplished with the extra tree and random forest classifier, and predicted results are accurate. This approach achieves good accuracy but takes large training times in packet level due to large amounts of data and unbalanced data.

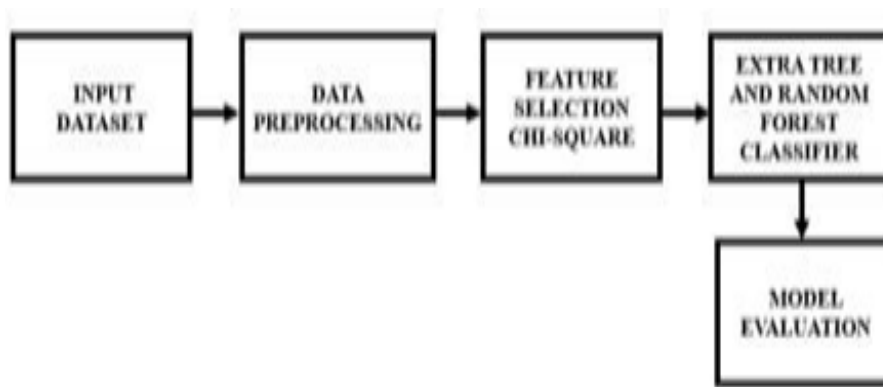


Figure 1: Block diagram of Proposed System



Article Title: IoT Network Traffic Classification Using Machine Learning Algorithms: An Experimental Analysis

5 Results and Discussion

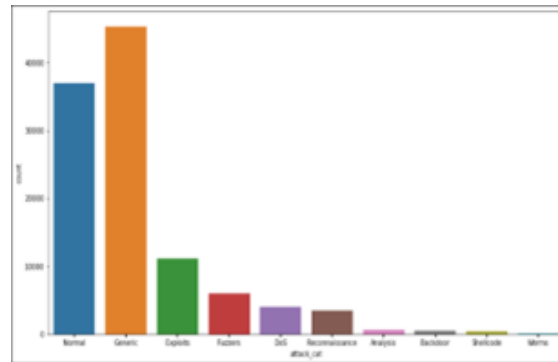


Figure 2: Attack-Cat

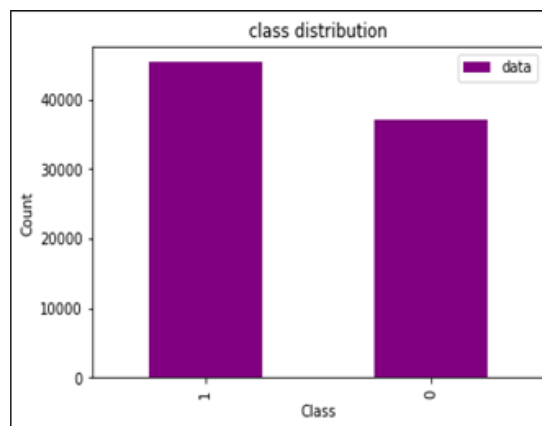
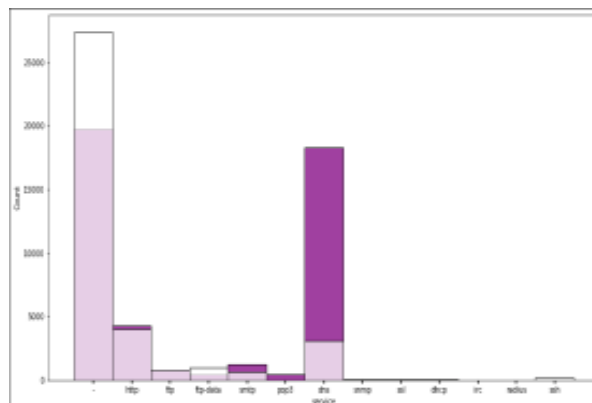


Figure 3: Class Distribution





Article Title: IoT Network Traffic Classification Using Machine Learning Algorithms: An Experimental Analysis

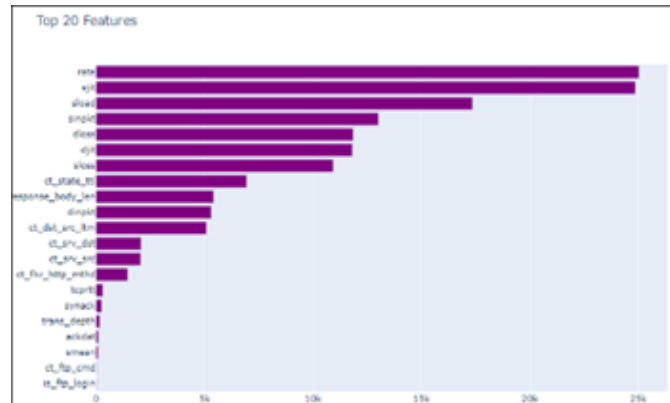


Figure 4: Top Feature

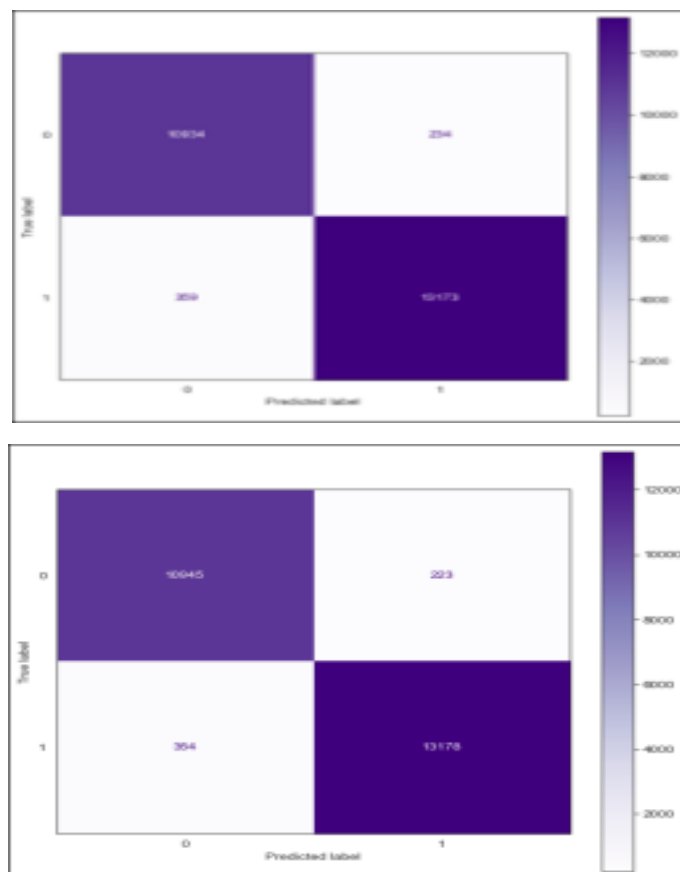


Figure 5: Performance Matrix for Extra Tree and Random Forest



Article Title: IoT Network Traffic Classification Using Machine Learning Algorithms: An Experimental Analysis

	Accuracy	Recall	Precision	F1-Score	time to train	time to predict	total time
ET	97.60%	97.60%	97.61%	97.60%	7.4	0.3	7.8
RF	97.60%	97.60%	97.67%	97.67%	7.9	0.2	8.2

Figure 6: Model Performance

6 Conclusion

This paper explains how to identify IoT devices and the network flows they generate using traffic analysis and two related techniques. The concept is straightforward and simple to implement in corporate settings. Furthermore, it has been established that common criteria for statistical classification of network traffic can also be utilised to identify trends and traffic classes. In the future, researchers intend to employ the IEEE-assigned name resolution and MAC standard to identify families of devices in real time and incorporate them, along with a traffic monitor.

7 Future Scope

In the future, researchers intend to employ the IEEE-assigned name resolution and MAC standard to identify families of devices in real time and incorporate them, along with a traffic monitor.

References

1. S. Egea; A. Rego Manez; B. Carro; A. Sánchez-Esguevillas; J. Lloret, Year: 2018, "Intelligent IoT Traffic Classification Using Novel Search Strategy for Fast-Based-Correlation Feature Selection in Industrial Environments," in IEEE Internet of Things Journal, Vol: 5, no: 3, pp. 1616-1624.
2. B. Mohammed et al., Year: 2020, "Edge Computing Intelligence Using Robust Feature Selection for Network Traffic Classification in Internet-of-Things," in IEEE Access, Vol: 8, pp. 224059-224070.
3. Umair M.B; Iqbal Z; Bilal M., Almohamad; T.A., Nebhen J; Mehmood R.M, Year: 2021, "An Efficient Internet Traffic Classification System Using DeepLearning for IoT".
4. Liu, D; Xu; X.; Liu; M; Liu, Y., Year: 2021, "Dynamic traffic classification algorithm and simulation of energy Internet of things based on machine learning. Neural Computing and Applications" Vol: 33, no: 9, pp. 3967-3976.
5. M. R. P. Santos; R. M. C. Andrade; D. G. Gomes; A. C. Callado, Year: 2018, "An efficient



Article Title: IoT Network Traffic Classification Using Machine Learning Algorithms: An Experimental Analysis

approach for device identification and traffic classification in IoT ecosystems," IEEE Symposium on Computers and Communications (ISCC) pp. 00304-00309.

6. Zhu, M.Y; Chen, Z; Chen, K.F; Lv, N; Zhong, Y., Year: 2022, "Attention-based federated incremental learning for traffic classification in the Internet of Things" Computer Communications, Vol: 185, pp. 168-175.

7. M R P Santos et al, Year: 2018, "An efficient approach for device identification and traffic classification in IoTecosystems".