README FILE

Introduction:

In this project, we analyze the performance of IoT attack detection using different dataset sizes. We use the machine learning algorithms to detect attacks on IoT devices and measure the accuracy, precision, f1 score and recall of these algorithms with different dataset sizes.

The main aim is to analyze the effect of dataset sizes on the accuracy of Machine Learning models. Also the ratio of the dataset means how imbalanced our dataset is and how it affects the accuracy.

Requirements:

To run this project, you need the following software and libraries installed on your system:

1. Python 3.x
2. Jupyter Notebook
3. Scikit-learn
4. Pandas
5. Numpy
6. Matplotlib
7. openpyxl

Dataset:

We use the attack (class type1 and class type 2) and non-attack datasets, which are generated through the net sim stimulator for intrusion detection in IoT devices. The different size datasets are 1:1 which is balanced, 1:2 is slightly imbalanced, and 1:3 which is an imbalanced dataset. Each of them contains 12 types of features and we use these dataset for our analysis.

Classification Models Used -

Total six Machine learning models have been used to train and analysis of our different datasets, these are: K-nearest neighbors algorithm, Decision Trees, Random Forest, Gradient Boosting, Naive Bayes and Support Vector Machine.

Project Structure:

The project has the following files:

1. `attack_detection_analysis.ipynb:` Jupyter notebook that contains the code for machine learning algorithms, data visualization, and performance analysis.
2. `.csv`: Dataset file in CSV format.
3. `README.md:` Readme file for the project.

Usage:

To run the project, follow these steps:

1. Clone the repository or download the project files to your local system.
2. Open the `attack_detection_analysis.ipynb` file in Jupyter Notebook.
3. Run the code cells in the notebook.
4. The notebook generates performance metrics and visualizations for machine learning algorithms with different dataset sizes.