How to Use the Code

Overview

This document explains how to run the provided code and use the datasets. The code focuses on anomaly detection using machine learning, along with network packet data and real-world datasets.

1. Required Tools and Dependencies

1. **Python (v3.8 or later)**
   - Install from Python.org.
2. **Python Libraries** Install these using `pip install <library_name>`:
   - `numpy`
   - `pandas`
   - `scikit-learn`
   - `tensorflow`
   - `scapy` (for `.pkt` files)
3. **Dataset Tools**
   - Extract the provided `datastreaming.zip` file.
   - Ensure you have a CSV viewer like Excel or any text editor.

2. How to Set Up the Environment

1. **Create a Folder**:
   - Place all the files (`twitter_training.csv`, `.pkt` files, `datastreaming.zip`) in one directory.
2. **Extract Dataset**:
   - Unzip `datastreaming.zip` into a folder using any zip extractor.
3. **Install Dependencies**:
   - Open a terminal or command prompt.

Run:
bash
Copy code
```
pip install numpy pandas scikit-learn tensorflow scapy
```

3. Steps to Run the Code

1. **Open the Notebook**:
   - Use Jupyter Notebook or Google Colab.
   - Open the `anomali.ipynb` file.
2. **Load Datasets**:
   - Ensure the datasets (e.g., `twitter_training.csv`) are in the same folder.
   - Update file paths in the notebook if necessary.

3. **Run Code Cells**:
    - Execute the notebook cells in order.
4. **Analyze Packet Files** (Optional):

Install `scapy` and use the following in a script:
python
Copy code

```
from scapy.all import rdpcap
packets = rdpcap('homeNet.pkt')
print(packets.summary())
```

4. Expected Outputs

- **Notebook Outputs**:
    - Metrics like accuracy, precision, recall, and F1 score for anomaly detection.
- **Packet Analysis**:
    - Network packet summaries (protocols, payload).

5. Common Issues

- **File Not Found**: Check if file paths in the code match the dataset locations.
- **Missing Libraries**: Install dependencies using `pip install`.
- **Permission Issues**: Run with administrator privileges if necessary.