Bot-IoT Dataset Cleaning Report

Raghad Alfarsi

1. Introduction

This report presents the data cleaning and preprocessing steps applied to the Bot-IoT dataset. The goal is to prepare a clean and structured dataset suitable for analysis and visualization.

2. Dataset Overview

The Bot-IoT dataset contains network traffic records labeled as either normal or malicious. It includes numerical and categorical features related to packet behavior and traffic statistics.

3. Data Cleaning and Preprocessing Steps

• **Removing Irrelevant Columns:** Columns such as IDs, timestamps, and MAC-related fields were removed as they do not contribute to meaningful analysis.

• **Encoding Categorical Features:** All categorical columns were converted into numerical values using label encoding to ensure compatibility with analytical tools.

• **Handling Missing Values:** Missing values in important columns were filled using the most frequent value, while remaining missing values were replaced with zero.

• **Feature Scaling:** Numerical features were standardized using StandardScaler to ensure consistent feature ranges.

4. Output Dataset

The final output is a cleaned dataset with no missing values, encoded categorical features, and scaled numerical features, ready for further analysis or visualization.

5. Conclusion

The preprocessing pipeline successfully transformed the raw Bot-IoT dataset into a clean and usable format. This ensures data quality and reliability for subsequent analytical tasks.