KNN Model on

**RT-IOT 2022 Dataset**

**Supervised By**

Dr. Eng. Muhammad Atta Khfagy

**Team Members**

Mohamed AbdElrahman Elkhlawy

Amina Younis Mohamed

Manar Hussain Mohamed

Norhan Nageh Elabd

Omar Mohamed Elhassan

Sara Nabil Kamel

Introduction

The RT-IoT2022 dataset is a proprietary dataset derived from a real-time IoT infrastructure. It serves as a comprehensive resource integrating a diverse range of IoT devices and sophisticated network attack methodologies.

This dataset encompasses both normal and adversarial network behaviors, providing a general representation of real-world scenarios. Incorporating data from IoT devices such as ThingSpeak-LED, Wipro-Bulb, and MQTT-Temp, as well as simulated attack scenarios involving Brute-Force SSH attacks, DDoS attacks using Hping and Slowloris, and Nmap patterns, RT-IoT2022 offers a detailed perspective on the complex nature of network traffic. The bidirectional attributes of network traffic are meticulously captured using the Zeek network monitoring tool and the Flowmeter plugin.

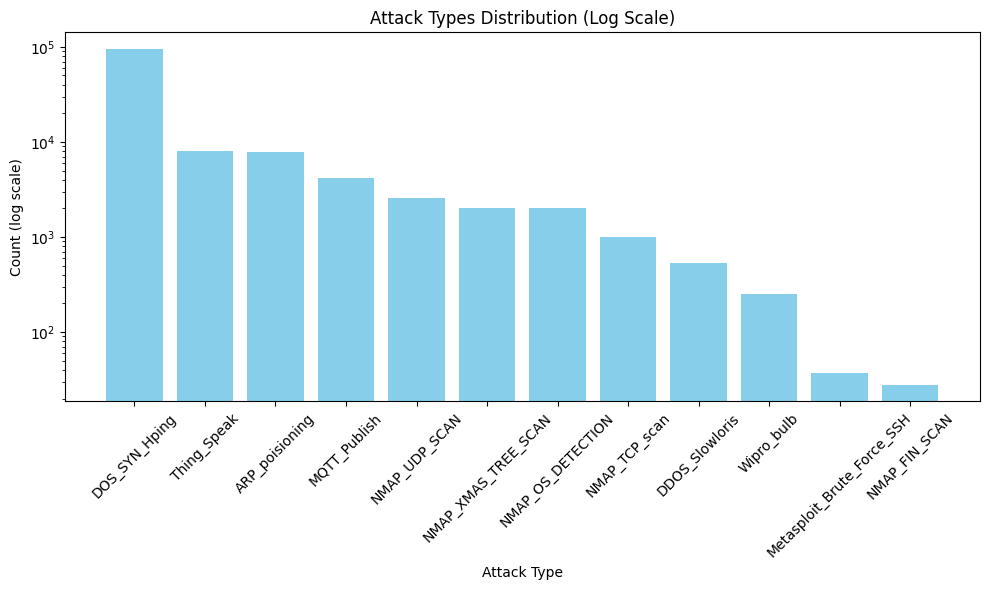
Researchers can leverage the RT-IoT2022 dataset to advance the capabilities of Intrusion Detection Systems (IDS), fostering the development of robust and adaptive security solutions for real-time IoT networks.

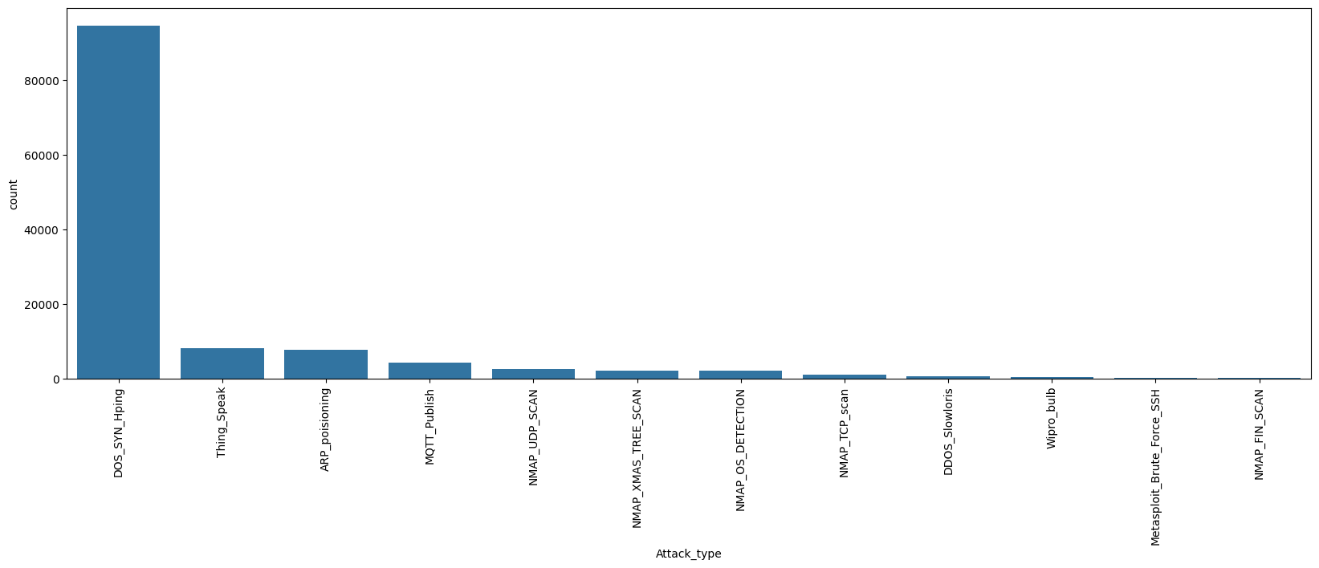
Step 1: Data Exploration

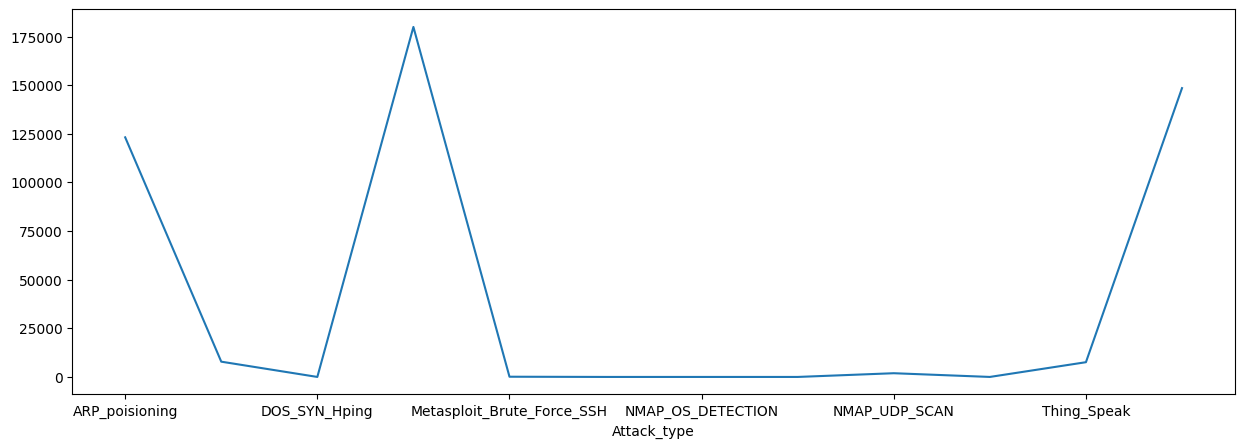
EDA Definition

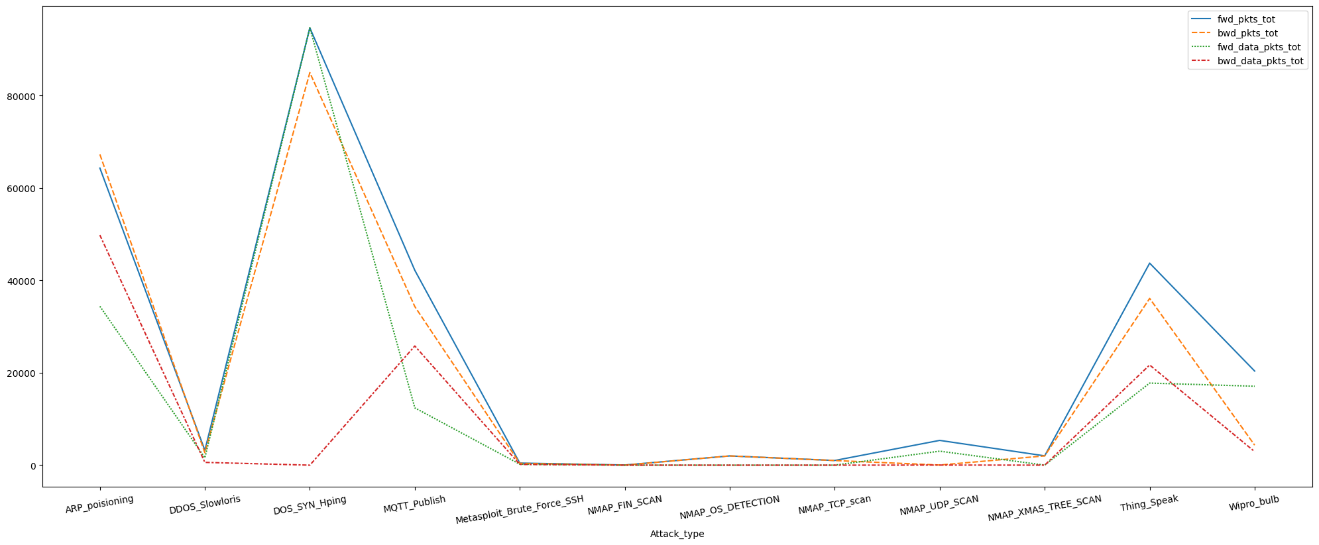
Exploratory Data Analysis (EDA) is a vital first step in building machine learning models. It involves using statistical methods and visualizations to understand the characteristics of your data. Through EDA, you can uncover patterns, identify relationships between features, and detect potential issues like missing values or outliers. These insights help prepare your data for modeling and can even inform the choice of machine learning algorithms best suited for the task.

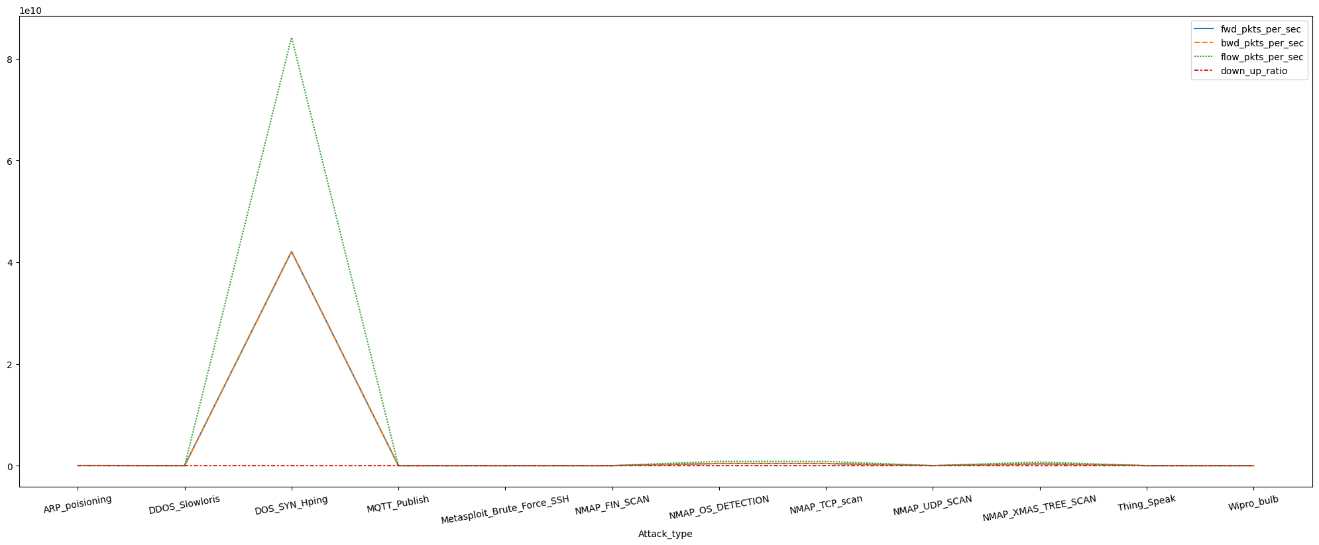
EDA for Model

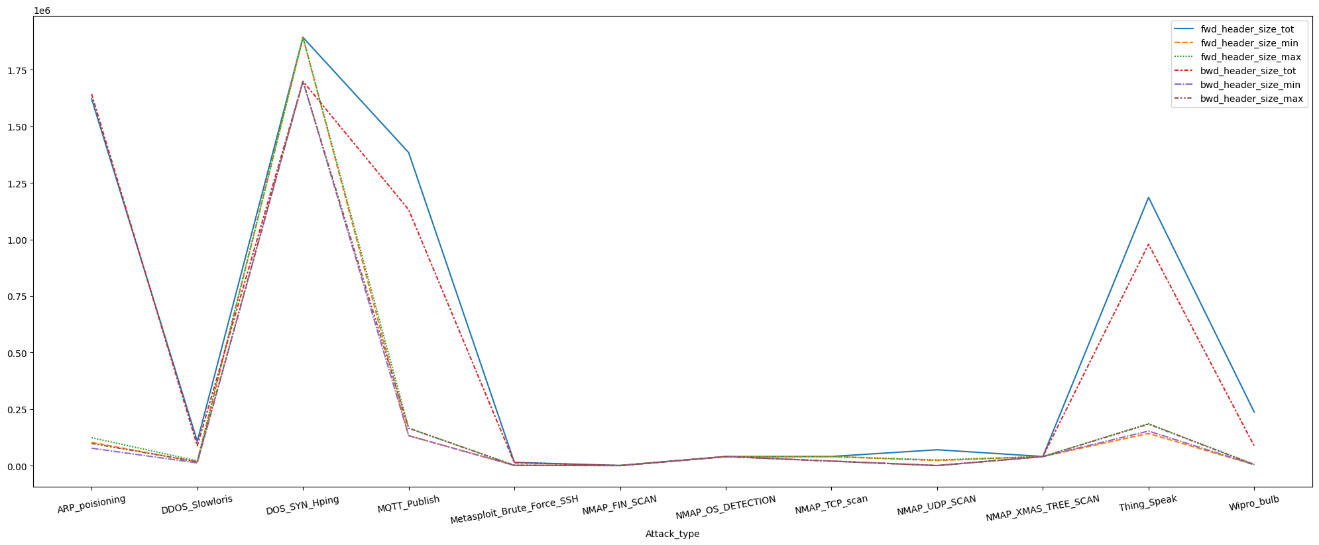
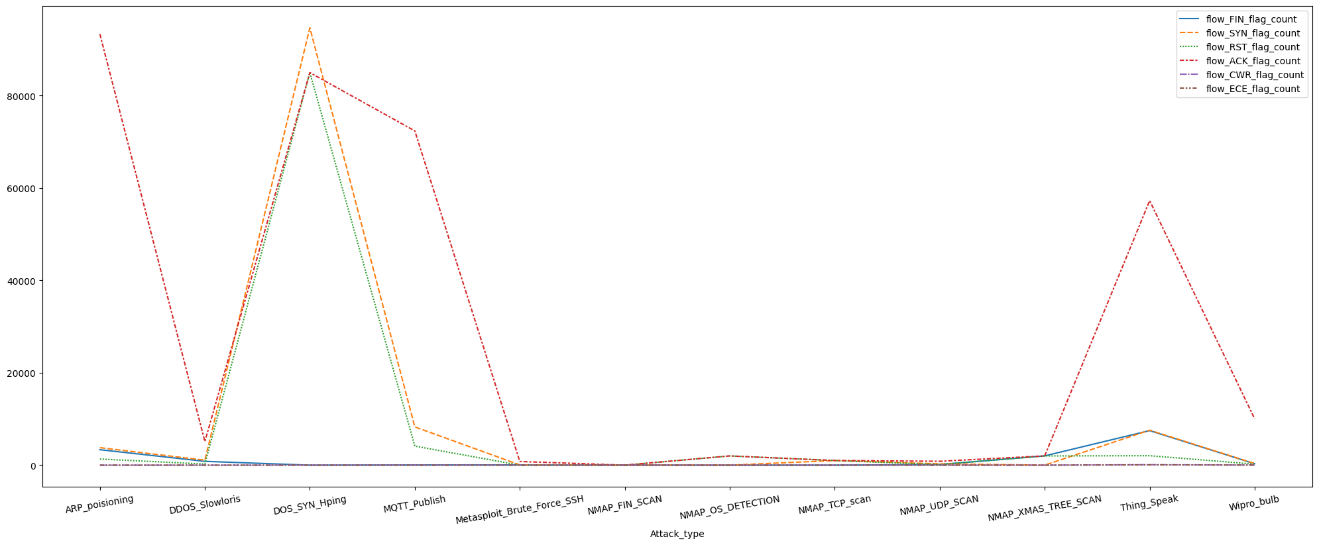


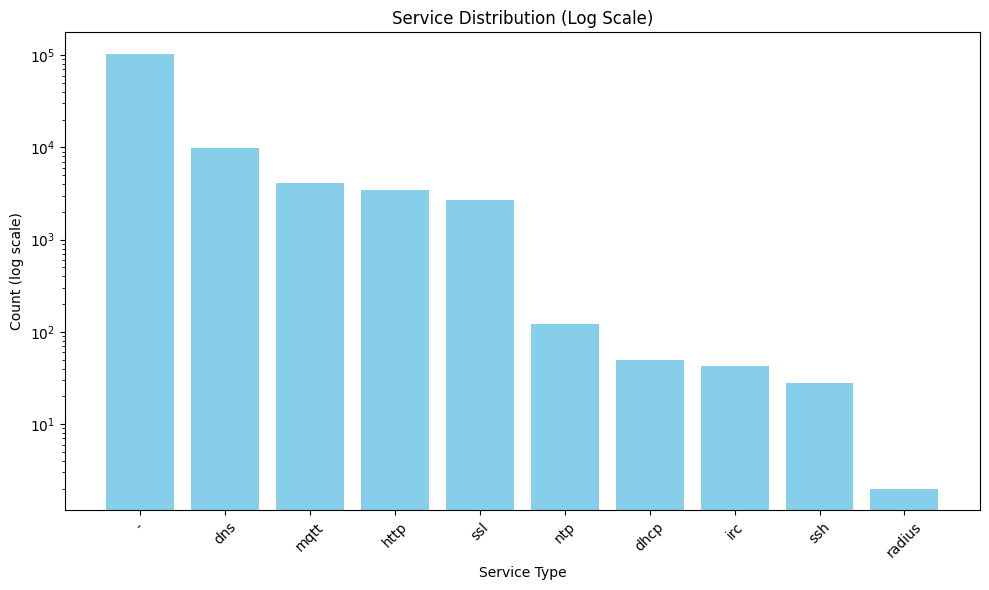
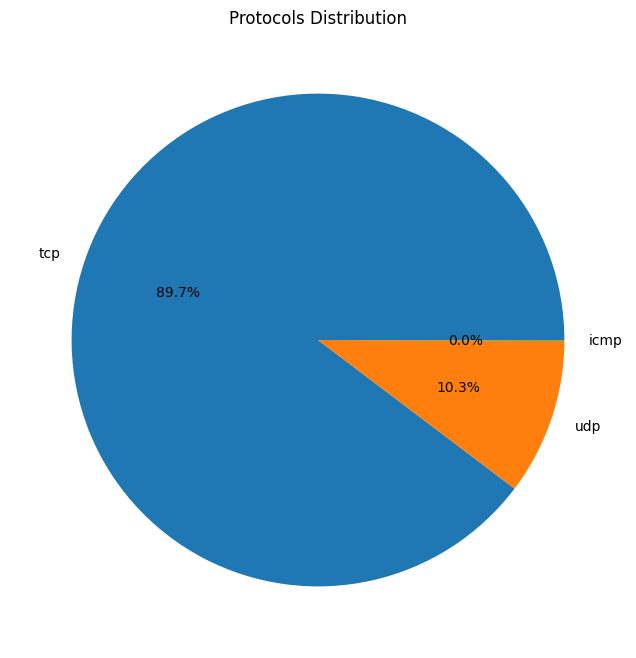
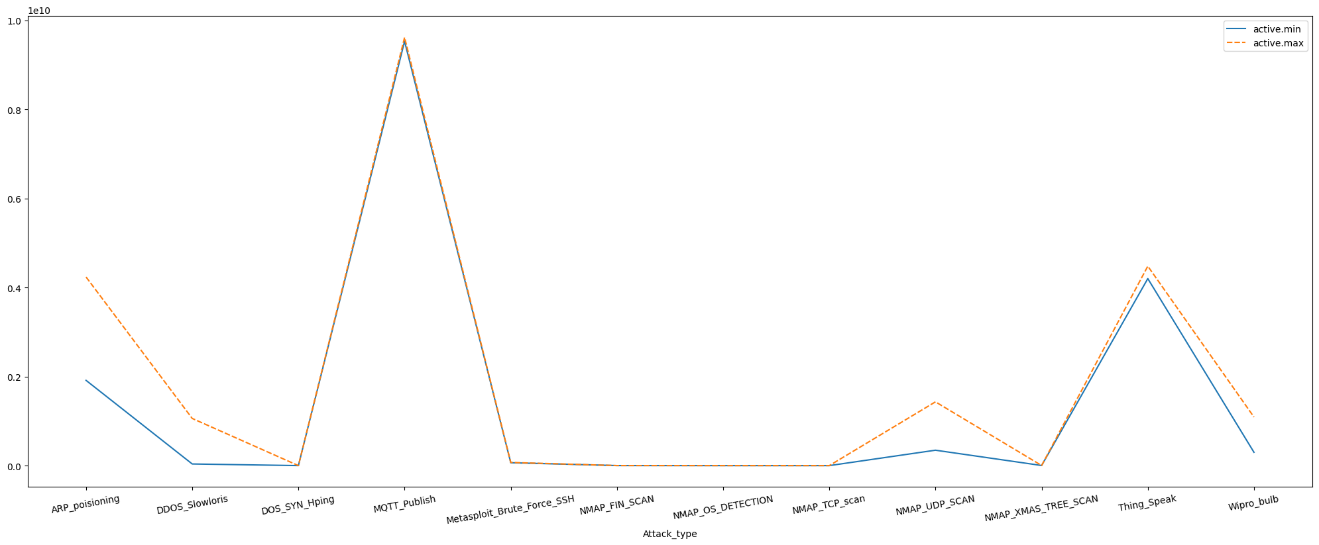
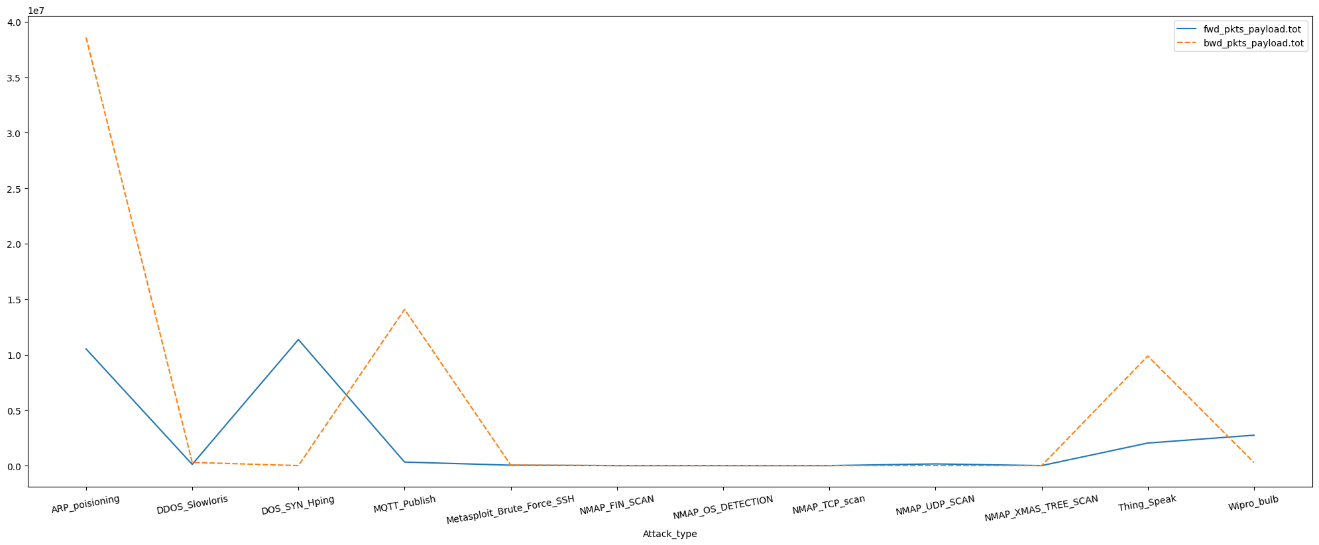
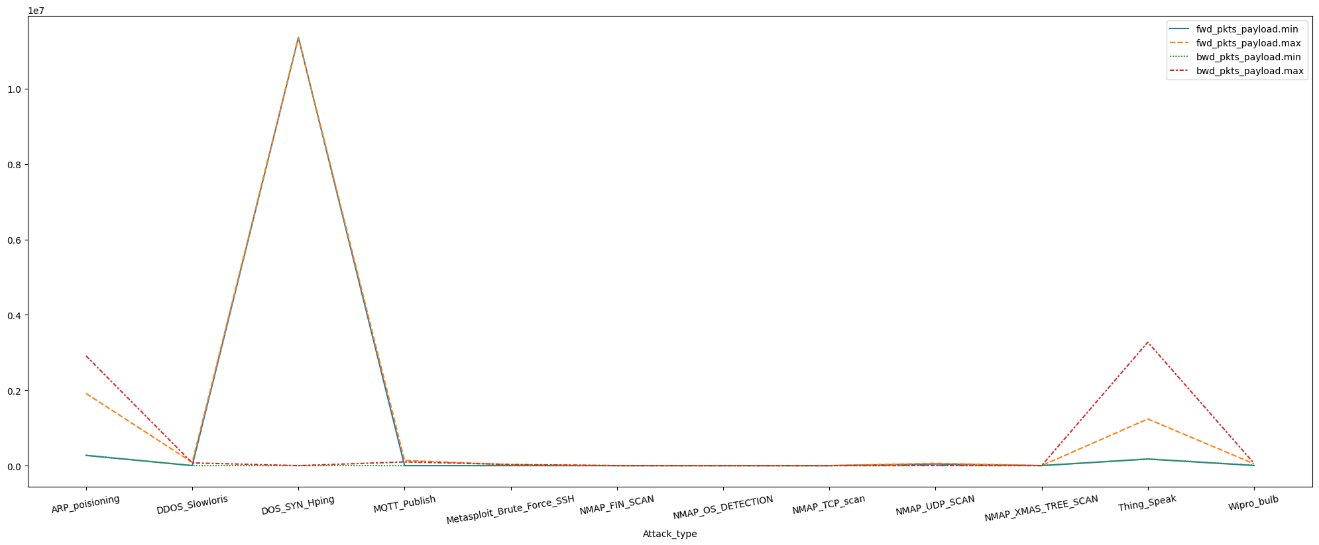
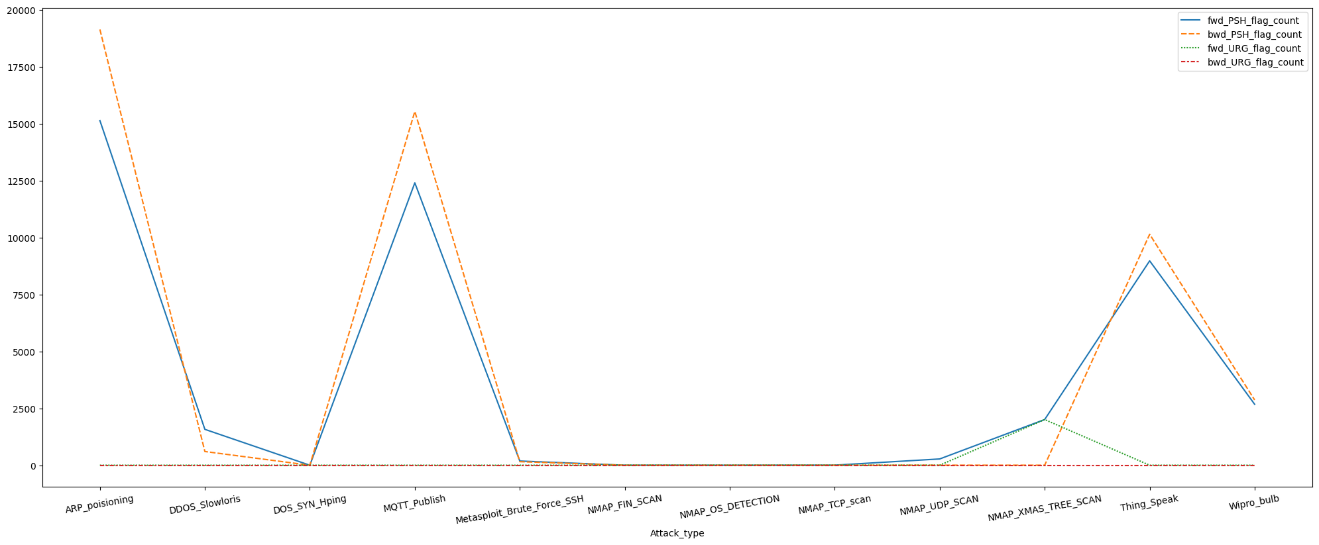










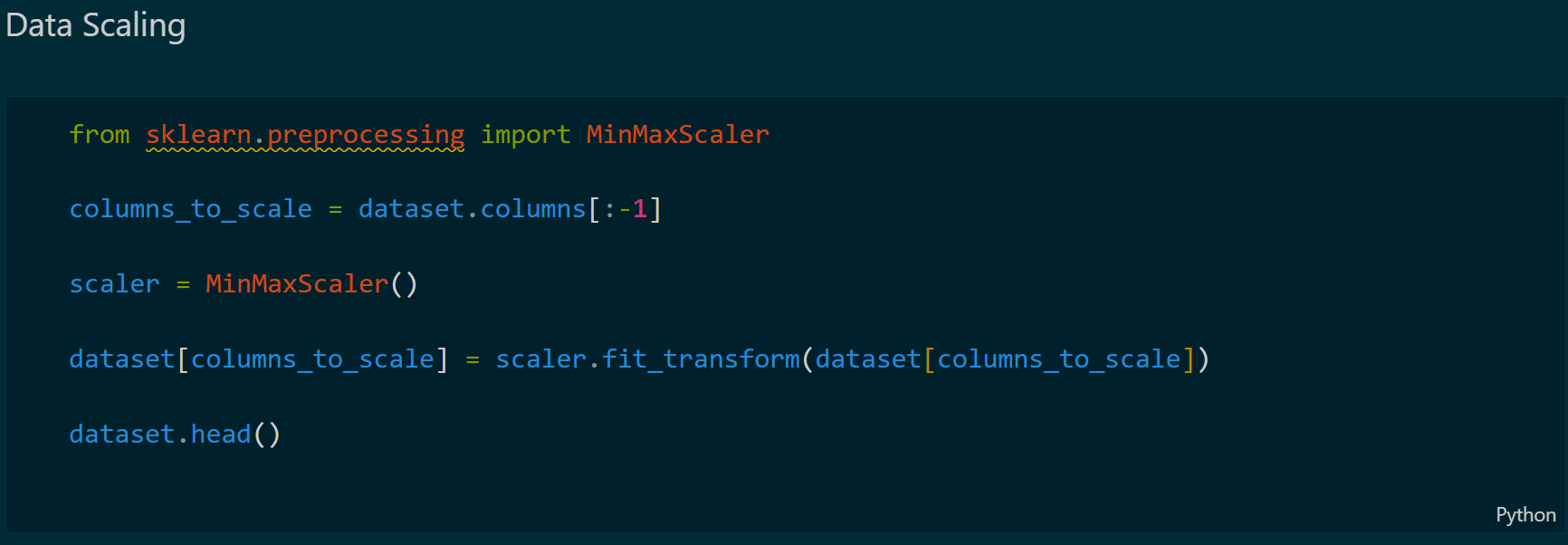


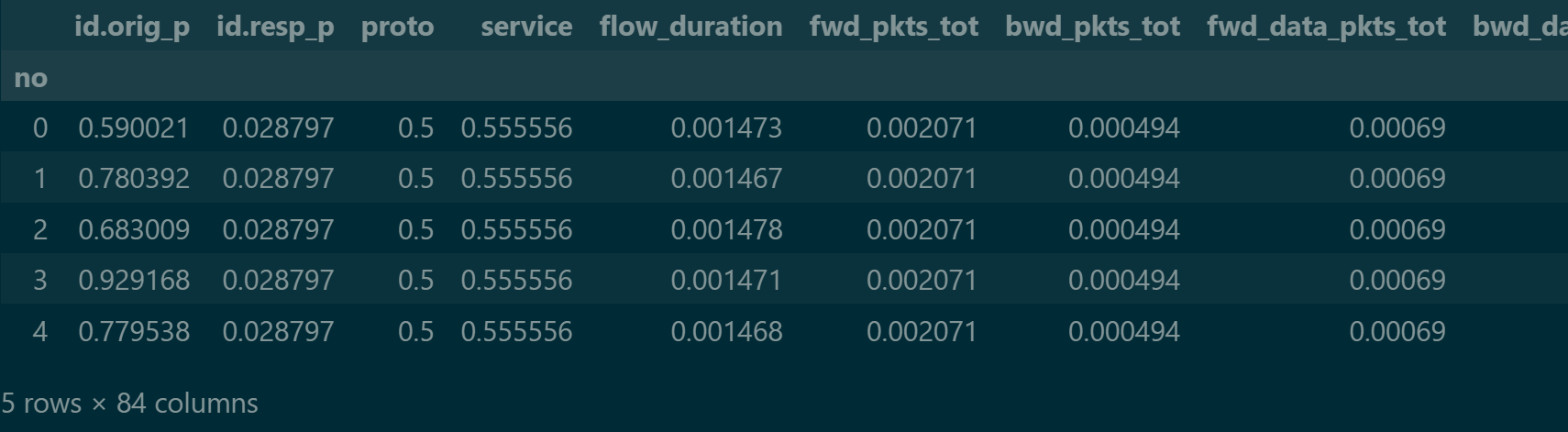
Step 2: Model Selection and Training

In this step we apply the following in the model:

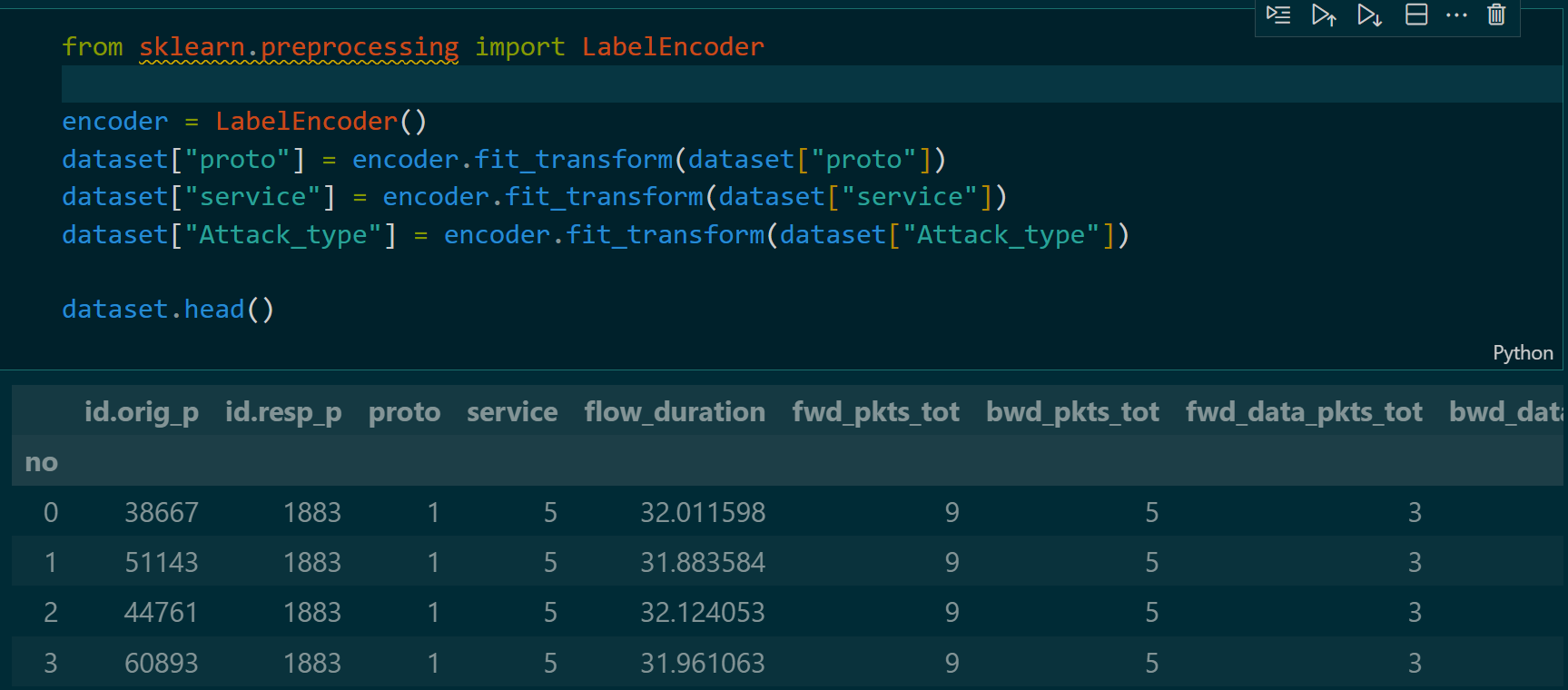
1. Label Encoding
2. Data Scaling
3. Split Dataset

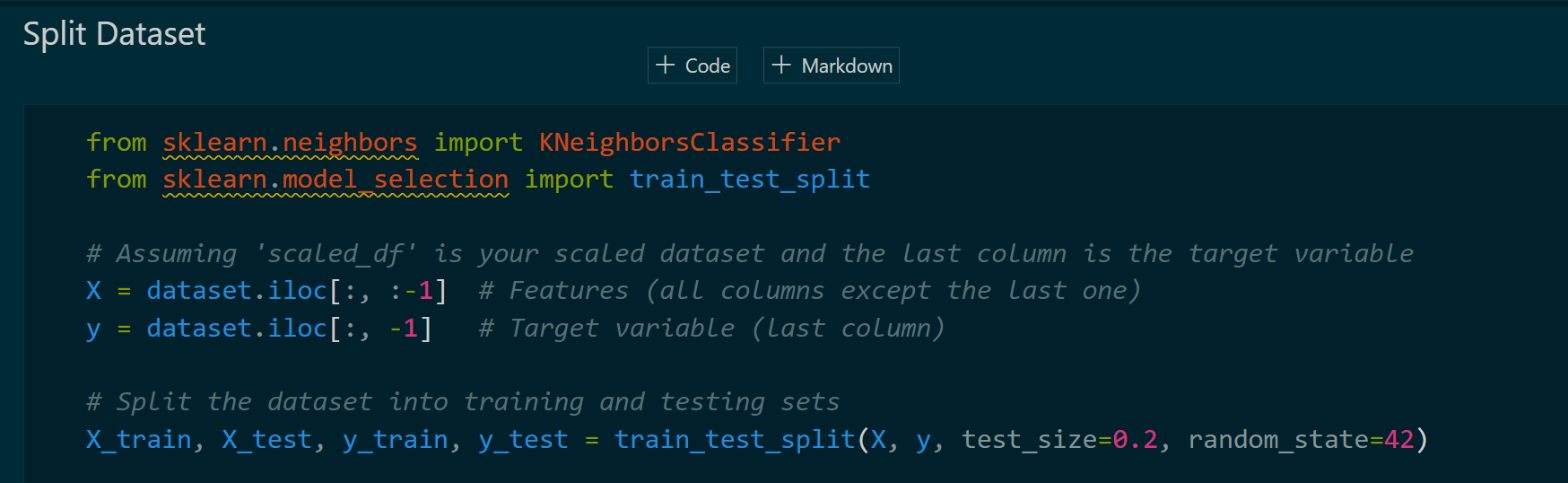
Figures

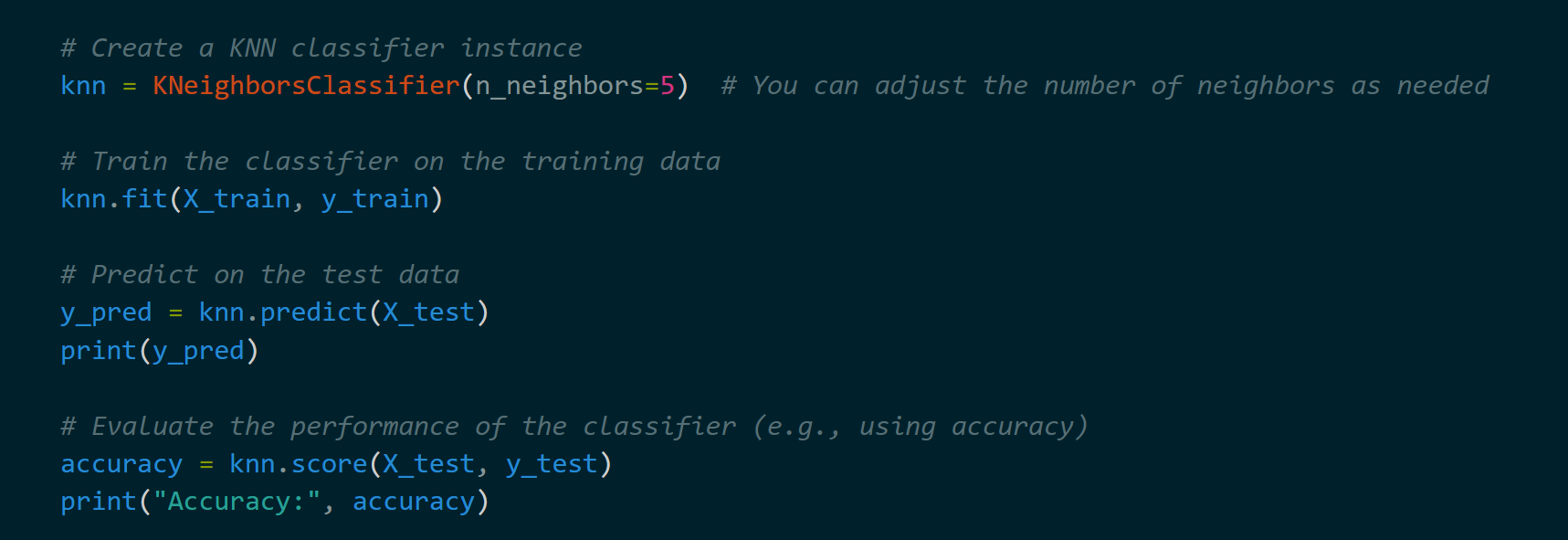










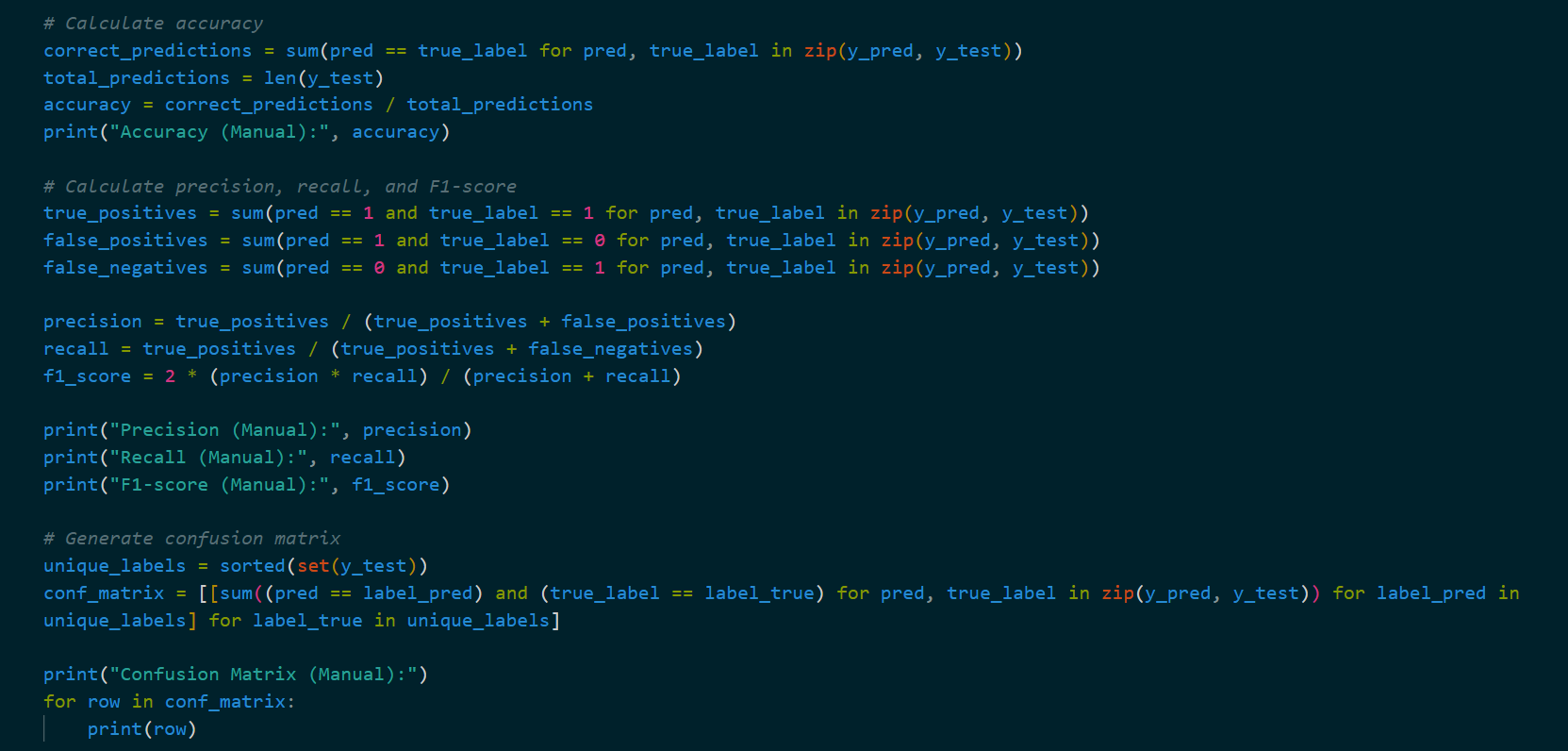


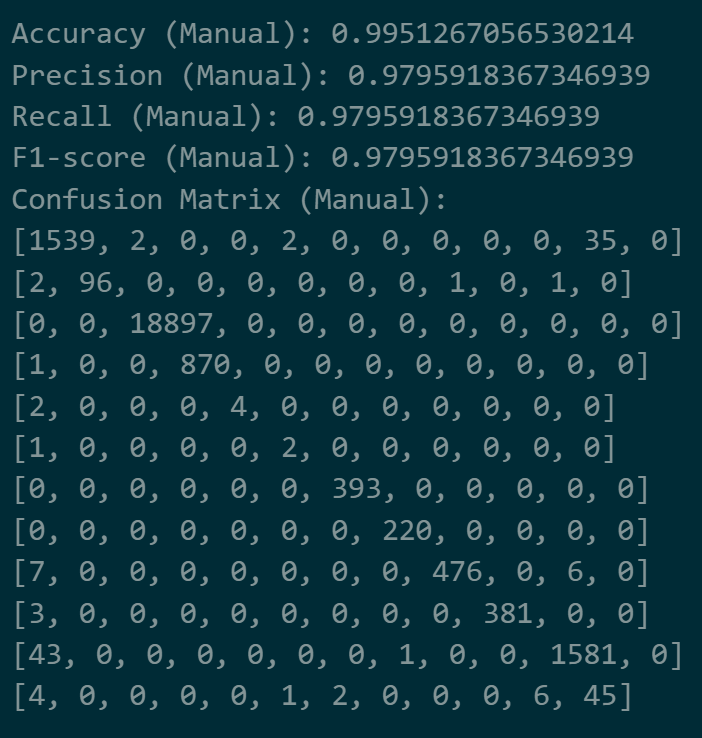


Step 3: Model Evaluation

In this step we calculate the model accuracy manually as following

Figures

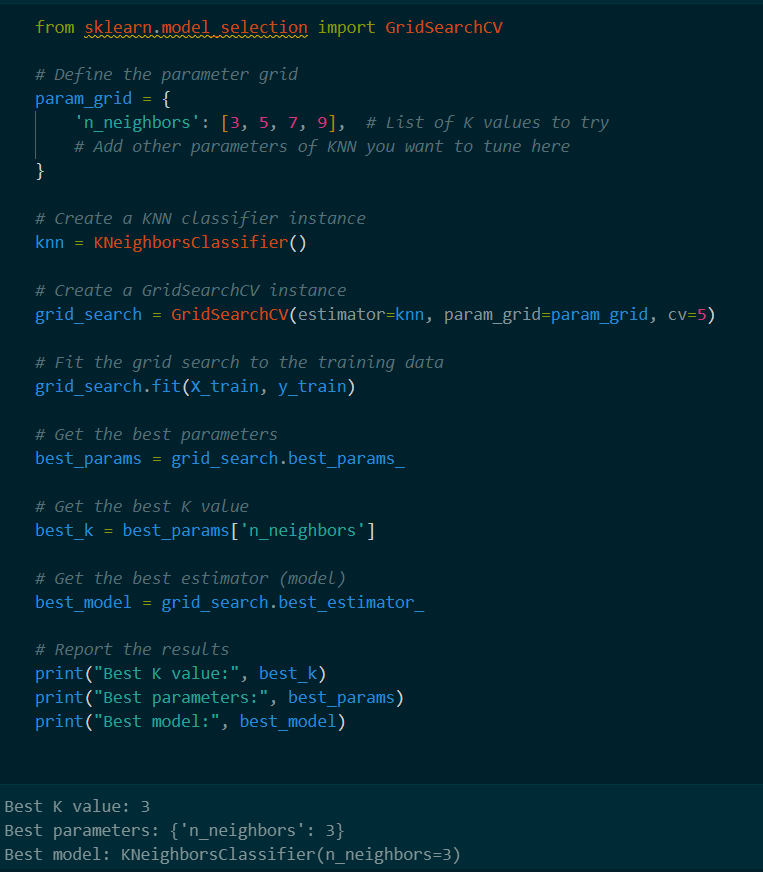




Step 4: Model Optimization

In this step we optimize the model to reach to the best number of K Neighbors as following

Figures



STep 5: Model Information

Google Colab URL

* <https://colab.research.google.com/drive/1J5R6SbhNmgJsDjEhwZdhMiLhLmAjPt_S>

**Thanks For Reading**