* Imported necessary libraries
* Read a DDoS dataset
* Removed white spaces before the column’s names
* Checked null values and remove it if exist
* Converted the labels to numerical which was not numerically
* Created histogram plot for each features
* Split dataset into training and testing set (70-30)
* Machine learning models training
* Testing the models
  + Random forest
    - Random Forest Metrics:
    - Accuracy: 0.9995
    - F1 Score: 0.9995
    - Precision: 1.0000
    - Recall: 0.9990
  + Removed less important features by feature engineering techniques.
  + Random forest after separated the top 45 important features.
    - Accuracy: 0.9994
    - F1 Score: 0.9994
    - Precision: 1.0000
    - Recall: 0.9989
    - ROC AUC Score: 1.00
  + Logistic Regression
    - Accuracy: 0.9339
    - F1 Score: 0.9404
    - Precision: 0.8970
    - Recall: 0.9882
    - ROC AUC Score: 0.99
  + Neural Network
    - Accuracy: 0.9855
    - F1 Score: 0.9863
    - Precision: 0.9824
    - Recall: 0.9903
    - ROC AUC Score: 0.99
* Hybrid Machine Learning Models training

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Precision | Recall | F1-score | Accuracy | Macro Avg | Weighted Avg |
| Stacking | ~0.9992 | ~1.0 | ~0.9996 | ~0.9996 | ~0.9992 | ~0.9996 |
| Boosting | ~0.9992 | ~1.0 | ~0.9996 | ~0.9996 | ~0.9992 | ~0.9996 |
| Voting | ~0.9992 | ~1.0 | ~0.9996 | ~0.9996 | ~0.9992 | ~0.9996 |

1. Introduced Explainable AI Techniques:
   * Attempted to use SHAP (SHapley Additive exPlanations) to interpret the model and understand feature contributions.

1. Data Preprocessing and Feature Engineering

1. Import necessary libraries
2. Read DDoS dataset
3. Remove white spaces from column names
4. Check and remove null values
5. Label Encoded
6. Remove less important features
7. Selected top 45 important features

2. Exploratory Data Analysis

1. Create histogram plots for each feature

3. Data Splitting

1. Split dataset into training (70%) and testing (30%) sets

4. Model Training and Testing

1. Random Forest

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

1. Logistic Regression

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

1. KNN

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

1. Support Vector Machine

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

1. Decision Tree

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

1. Neural Network

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

5. Hybrid Machine Learning Model 1 Training

* Base Model: LR, RF, SVM

1. Stacking

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

1. Boosting

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

1. Voting

* Train model
* Evaluate: Accuracy, F1 Score, Precision, Recall, Confusion matrix

6. Hybrid Machine Learning Model 2 Training

* Base Model: Gradient Boosting , XGBoost

1. Stacked Model (Gradient Boosting + XGBoost)

6. Explainable AI Techniques

|--> Apply SHAP to interpret model and understand feature contributions