TITLE : EXPLAINABLE AI

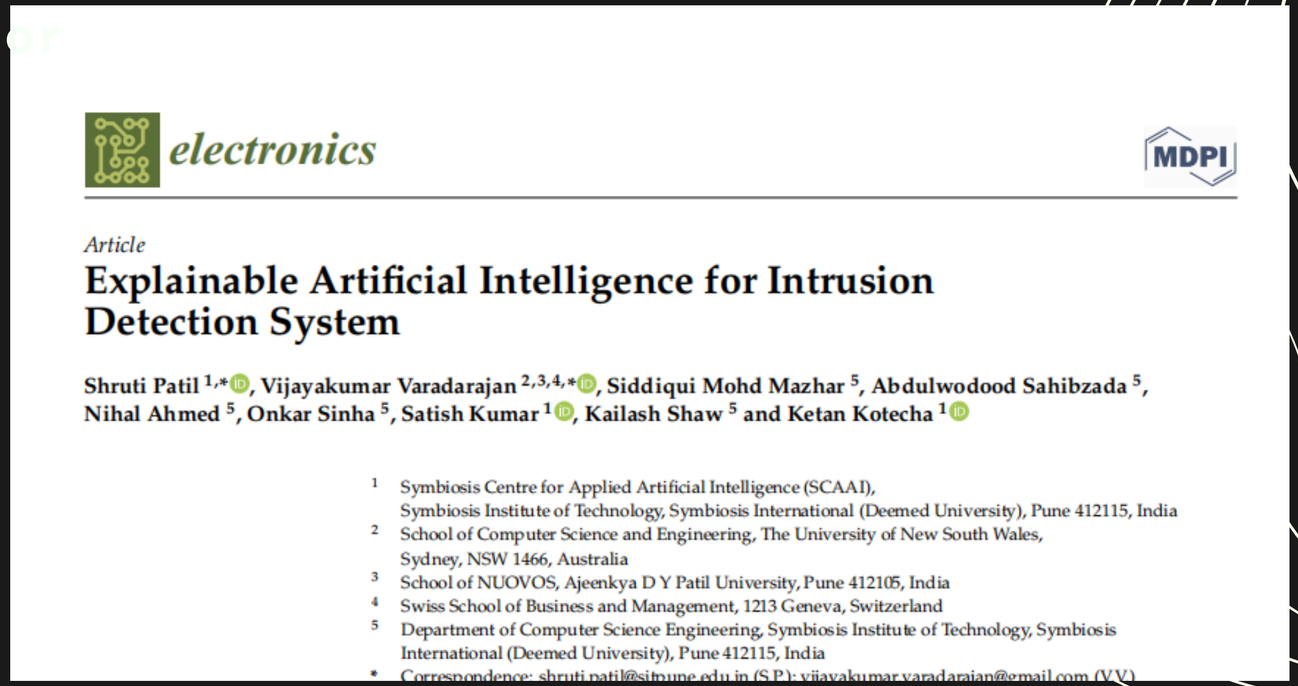
(ML PROJECT)

MANSURAH M - 22BAI1338
GOWTHAMI S - 22BAI1457

RESEARCH PAPER

Explainable Artificial Intelligence for Intrusion Detection System

Authors:
Shruti Patil,Vijayakumar Varadarajan,Siddiqui Mohd Mazhar,
Submission received: 18 July 2022 /
Revised: 14 September 2022 /
Accepted: 20 September 2022 /
Published: 27 September 2022



electronics

MDPI

Article

Explainable Artificial Intelligence for Intrusion Detection System

Shruti Patil [1,*], Vijayakumar Varadarajan [2,3,4,*], Siddiqui Mohd Mazhar [5], Abdulwodood Sahibzada [5], Nihal Ahmed [5], Onkar Sinha [5], Satish Kumar [1], Kailash Shaw [5] and Ketan Kotecha [1]

1 Symbiosis Centre for Applied Artificial Intelligence (SCAAI), Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India
2 School of Computer Science and Engineering, The University of New South Wales, Sydney, NSW 1466, Australia
3 School of NUOVOS, Ajeenkya D Y Patil University, Pune 412105, India
4 Swiss School of Business and Management, 1213 Geneva, Switzerland
5 Department of Computer Science Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India
* Correspondence: shruti.patil@sitpune.edu.in (S.P.); vijayakumar.varadarajan@gmail.com (V.V.)

OBJECTIVE

- The model incorporates the XAI algorithm LIME for better explainability and understanding of the black-box approach to intrusion detection system by observing in decision tree, random forest and support vector machine.
- The proposed IDS incorporates features from the CICIDS-2018 dataset .

WORKFLOW OF THE PAPER:

Workflow :

1. Data Loading
2. Data preprocessing
3. Feature Selection
4. ML model training
5. Applying LIME to the ML models (Decision tree, SVM, Random Forest classifier)
6. Generating user-understandable explanations

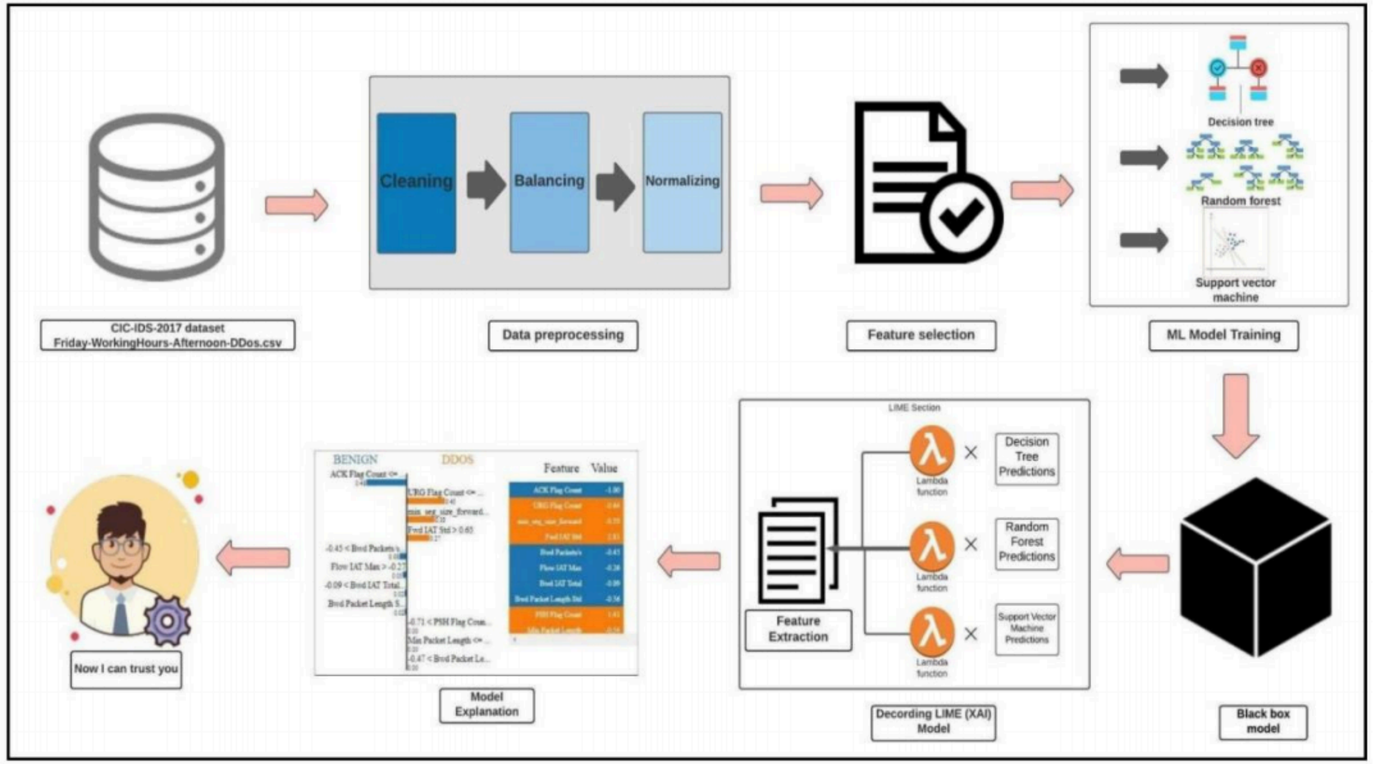
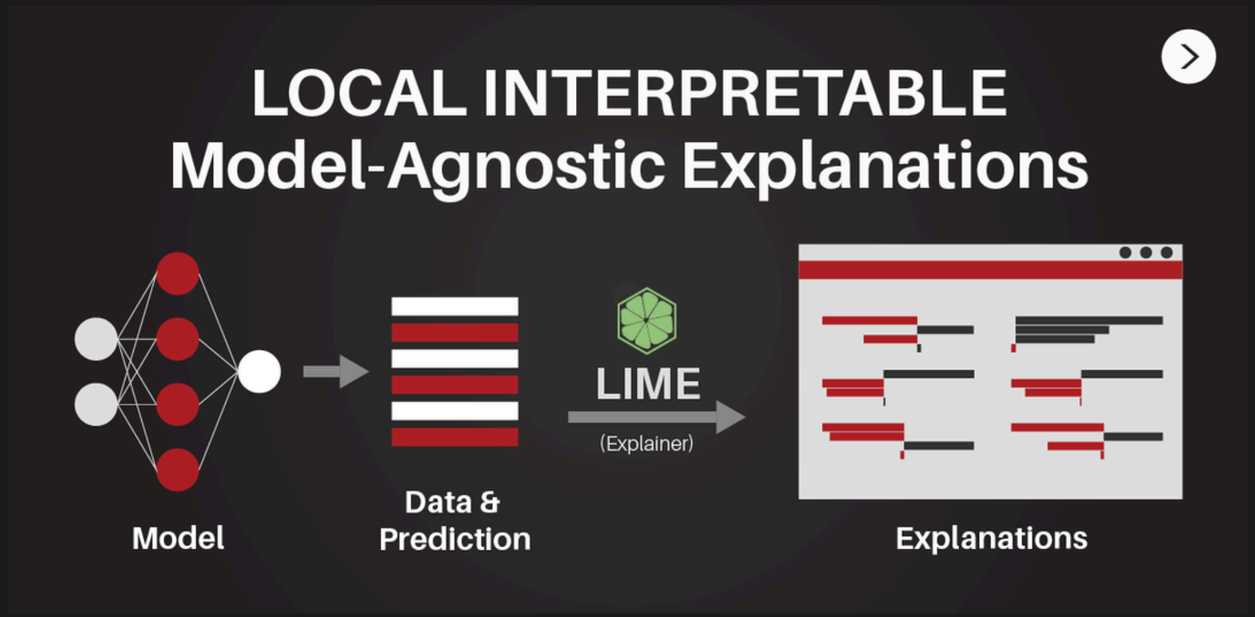


Figure 7. Explainable AI Framework.

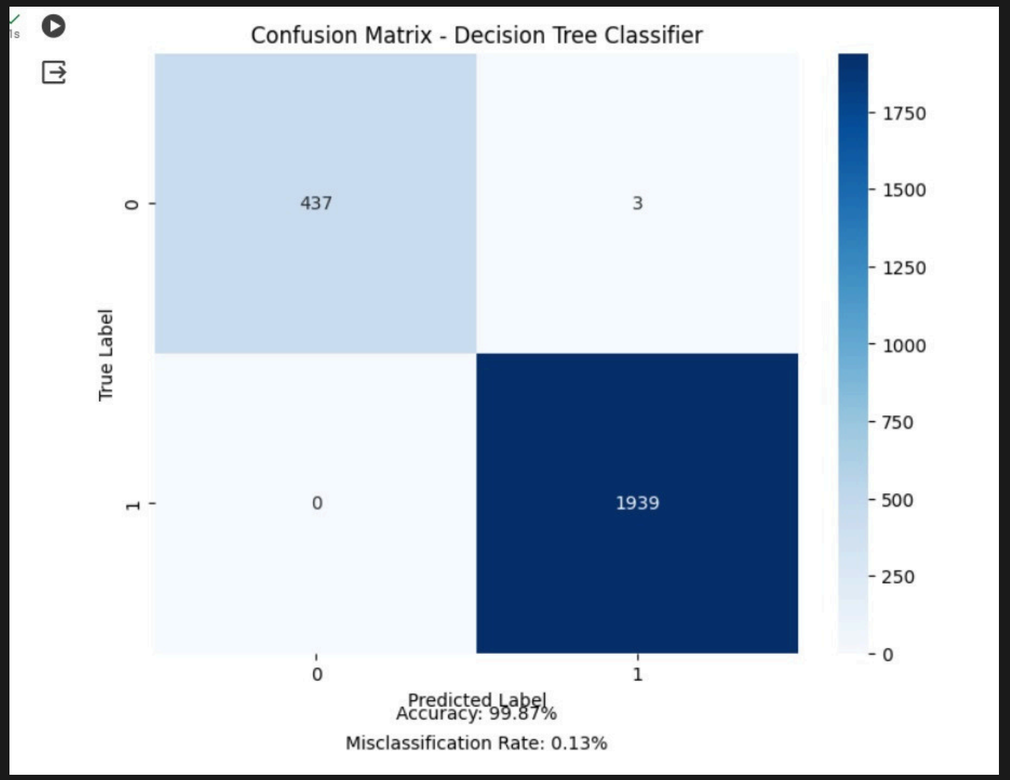
LIME

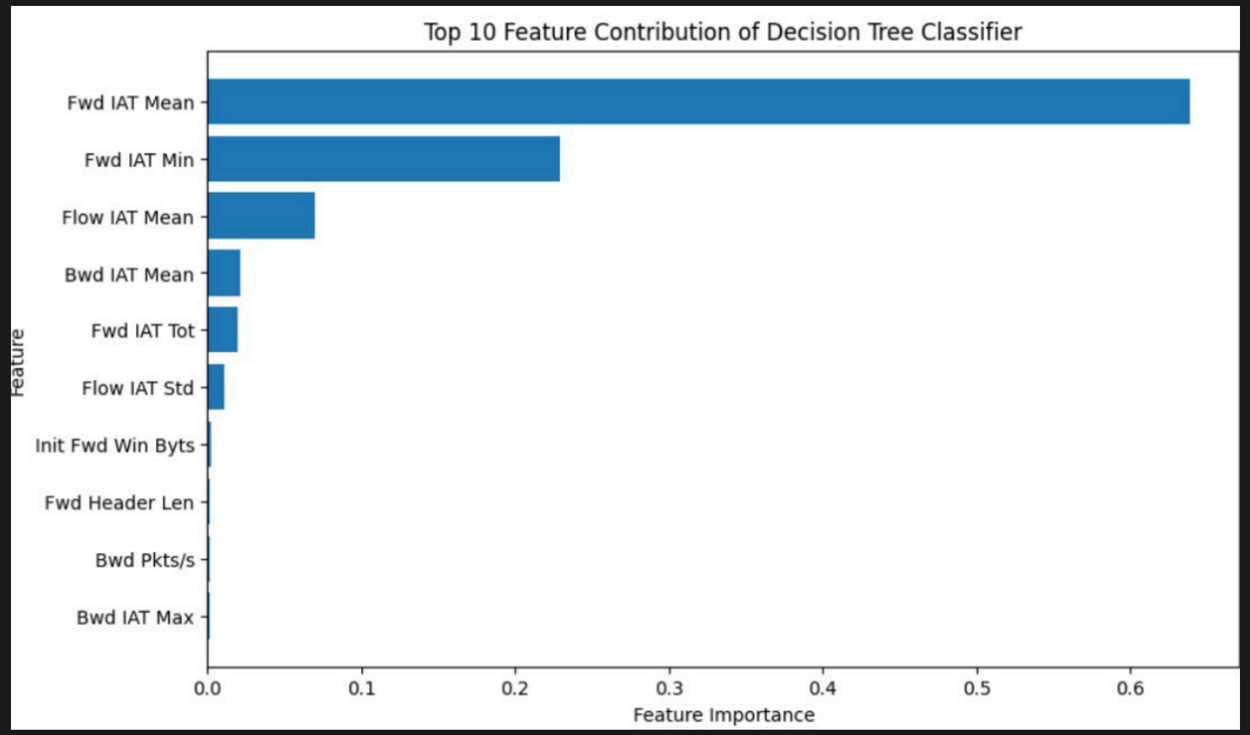


- LIME, Local Interpretable Model-agnostic Explanations, is a technique that can be used to explain the predictions made by any black-box classifier.

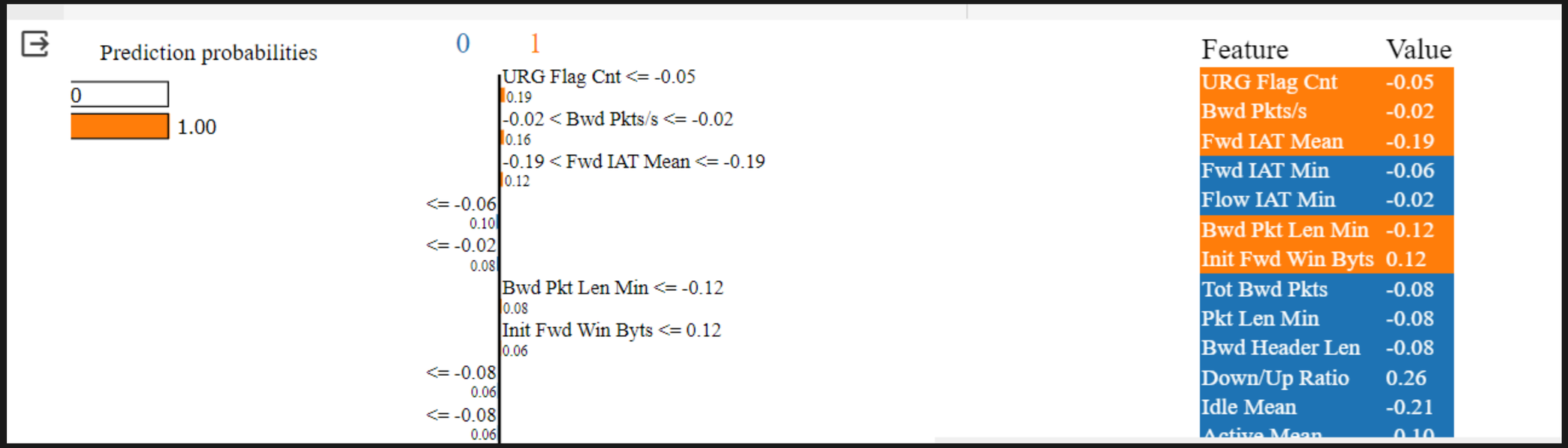
- The proposed IDS incorporates features from the CICIDS-2018 dataset.

DECISION TREE

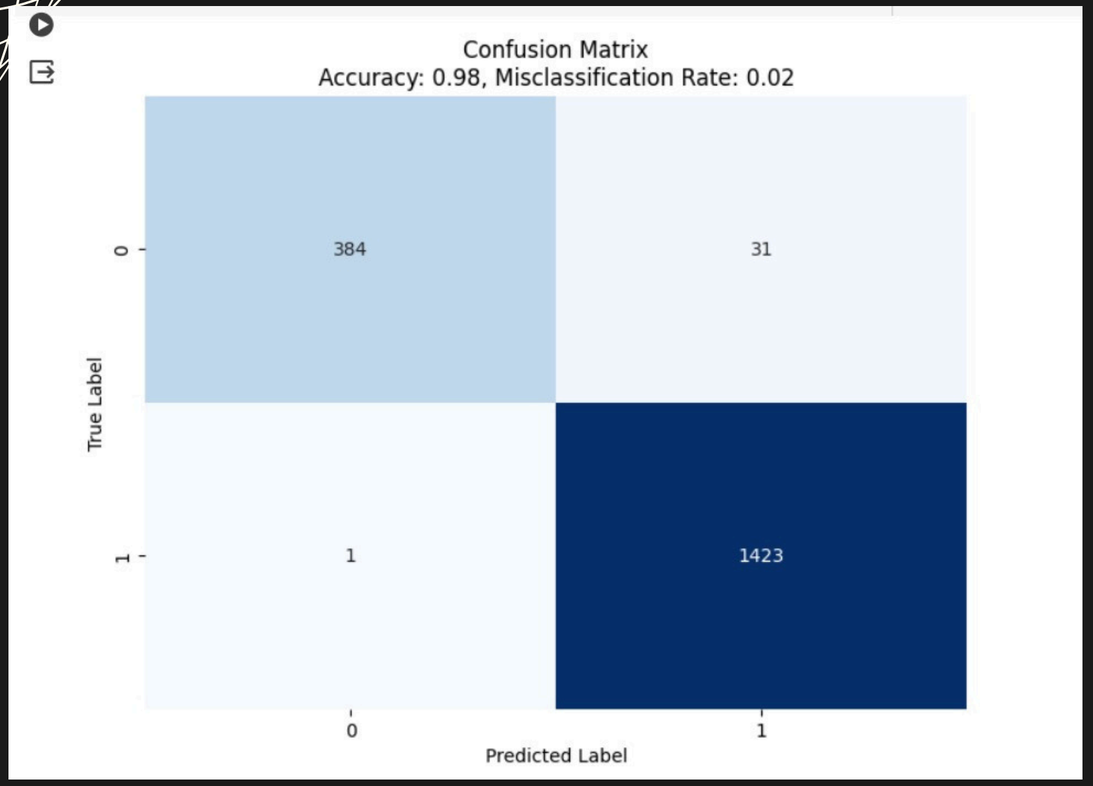


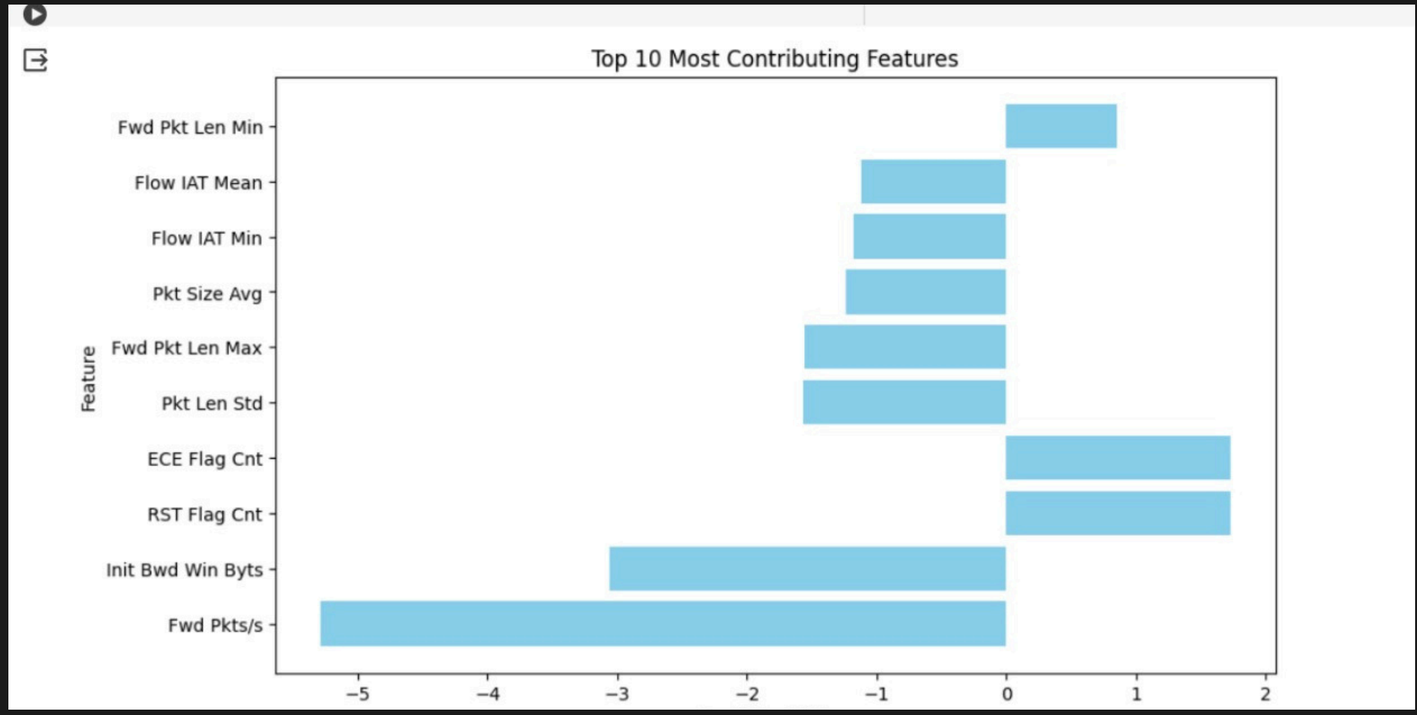


GENERATING LIME EXPLANATIONS FOR DECISION TREE

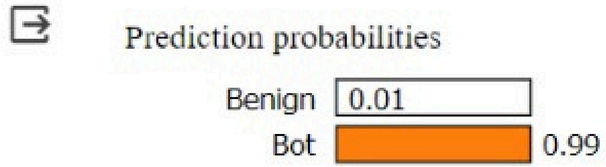


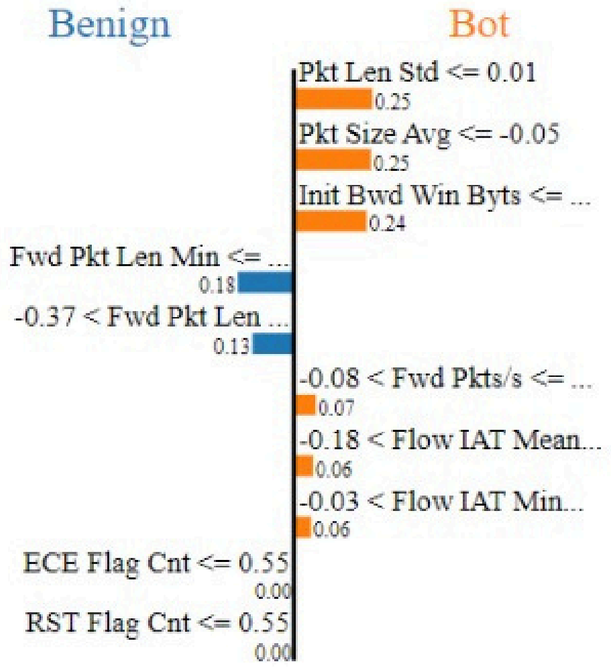
SVM


Confusion Matrix
Accuracy: 0.98, Misclassification Rate: 0.02
384
31
1
1423
True Label
Predicted Label
0
1
0
1


Top 10 Most Contributing Features
Fwd Pkt Len Min
Flow IAT Mean
Flow IAT Min
Pkt Size Avg
Fwd Pkt Len Max
Pkt Len Std
ECE Flag Cnt
RST Flag Cnt
Init Bwd Win Byts
Fwd Pkts/s
Feature
−5
−4
−3
−2
−1
0
1
2

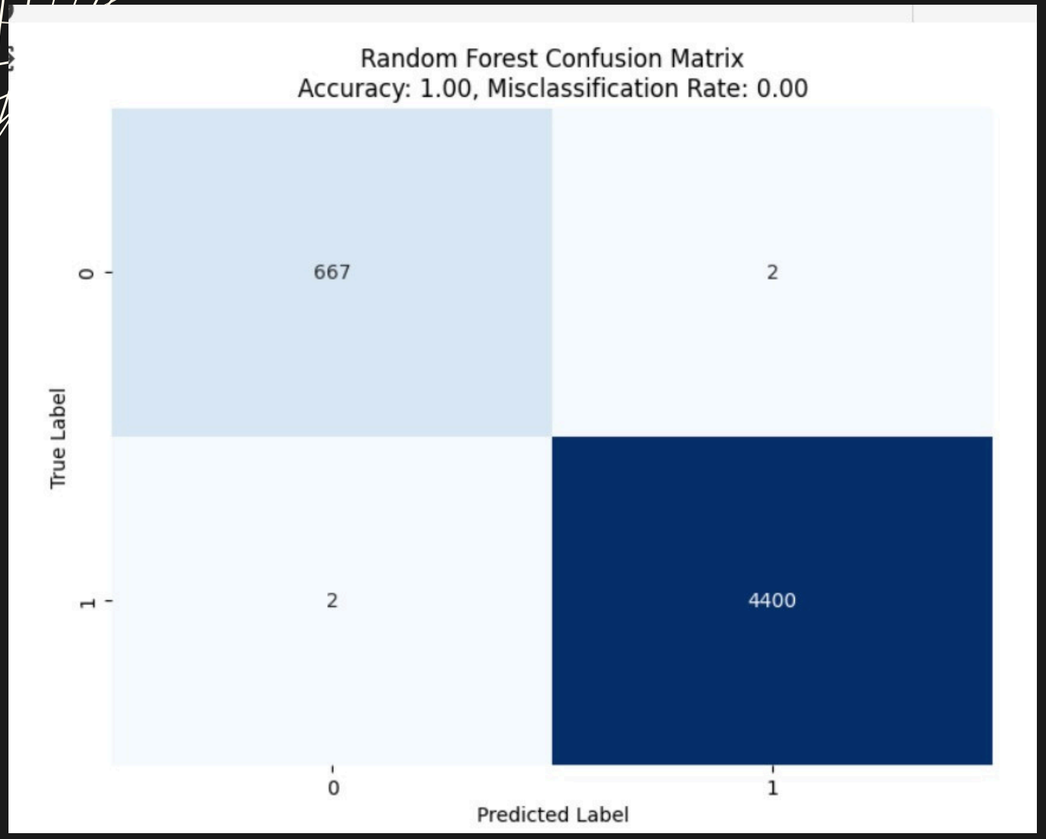
SVM

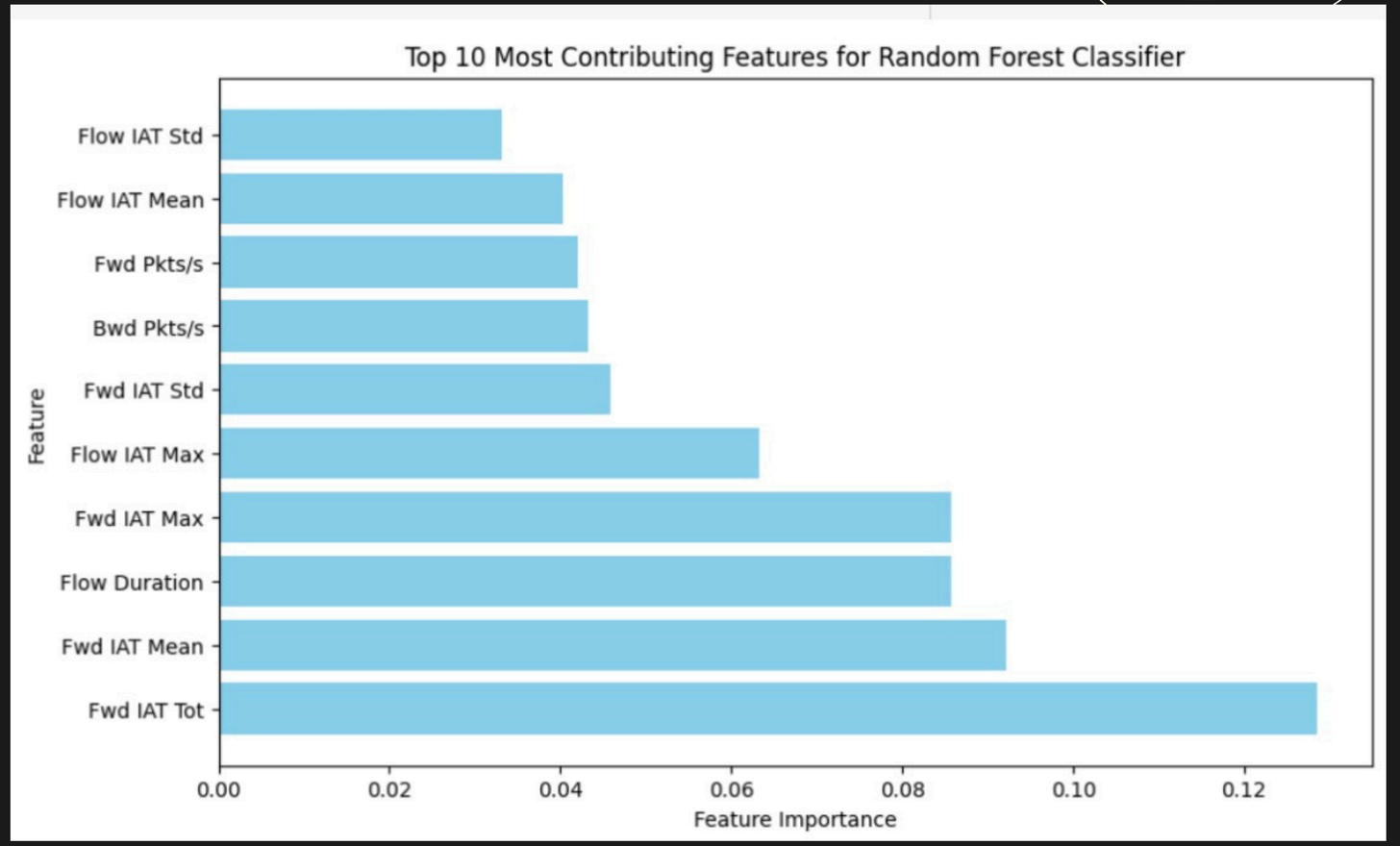




Feature	Value
Pkt Len Std	0.01
Pkt Size Avg	-0.05
Init Bwd Win Byts	-0.17
Fwd Pkt Len Min	-0.10
Fwd Pkt Len Max	0.32
Fwd Pkts/s	-0.08
Flow IAT Mean	-0.18
Flow IAT Min	-0.03
ECE Flag Cnt	0.55
RST Flag Cnt	0.55

RANDOM FOREST CLASSIFIER

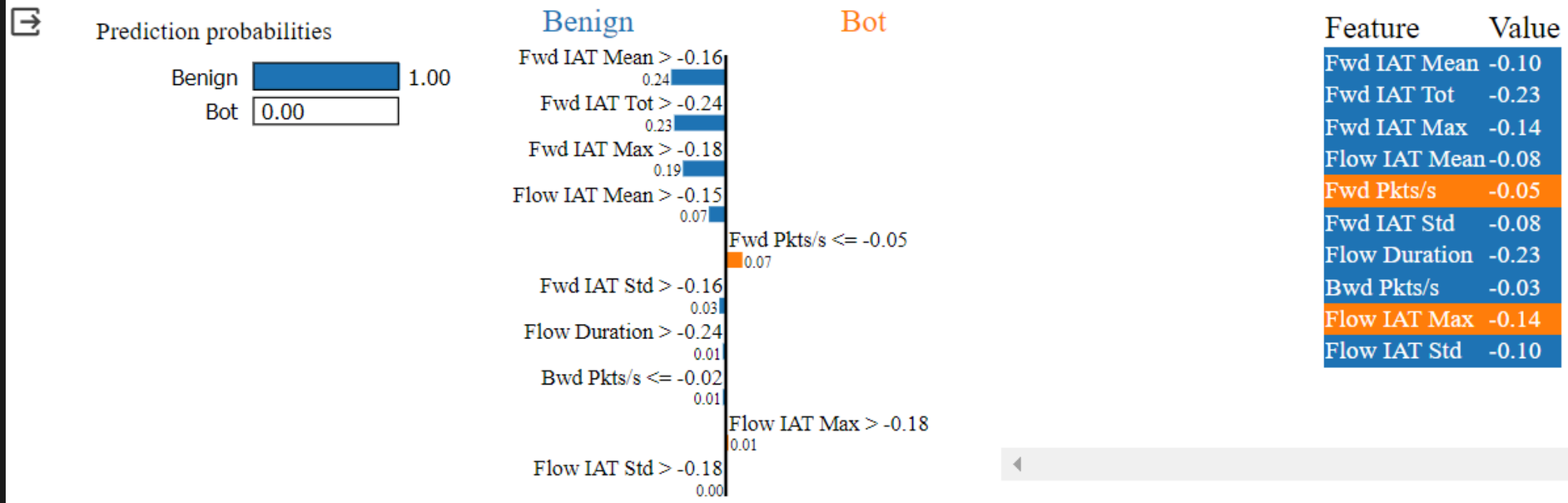




RANDOM FOREST CLASSIFIER

```
                                                    rf_classifier_top10.predict_proba,
                                                    num_features=len(top10_features_rf))


# Print LIME explanation
rf_explanation_top10.show_in_notebook()
```



THANK YOU