CERIAS

The Center for Education and Research in Information Assurance and Security

Ensemble Feature Selection for Network Intrusion Detection Systems Using Explainable AI: A Frequency-Based Approach

Ismail Bibers[1], and Mustafa Abdallah[1]

[1]Purdue University Indianapolis, Indiana, USA

ibibers@purdue.edu, abdalla0@purdue.edu

Introduction

- The research addresses the challenge of high-dimensional network traffic in **Intrusion Detection Systems (IDS)** and the need for efficient feature selection.
- It utilizes **explainable AI (XAI) methods** such as **SHAP**, Leave-One-Covariate-Out (**LOCO**), Profiled Weighting(**ProfWeight**), Permutation Feature Importance (**PFI**), and **DALEX** to improve interpretability and feature importance ranking.
- The study introduced **frequency-based ensemble approach** to identify the most critical features, enhancing detection accuracy and computational efficiency.

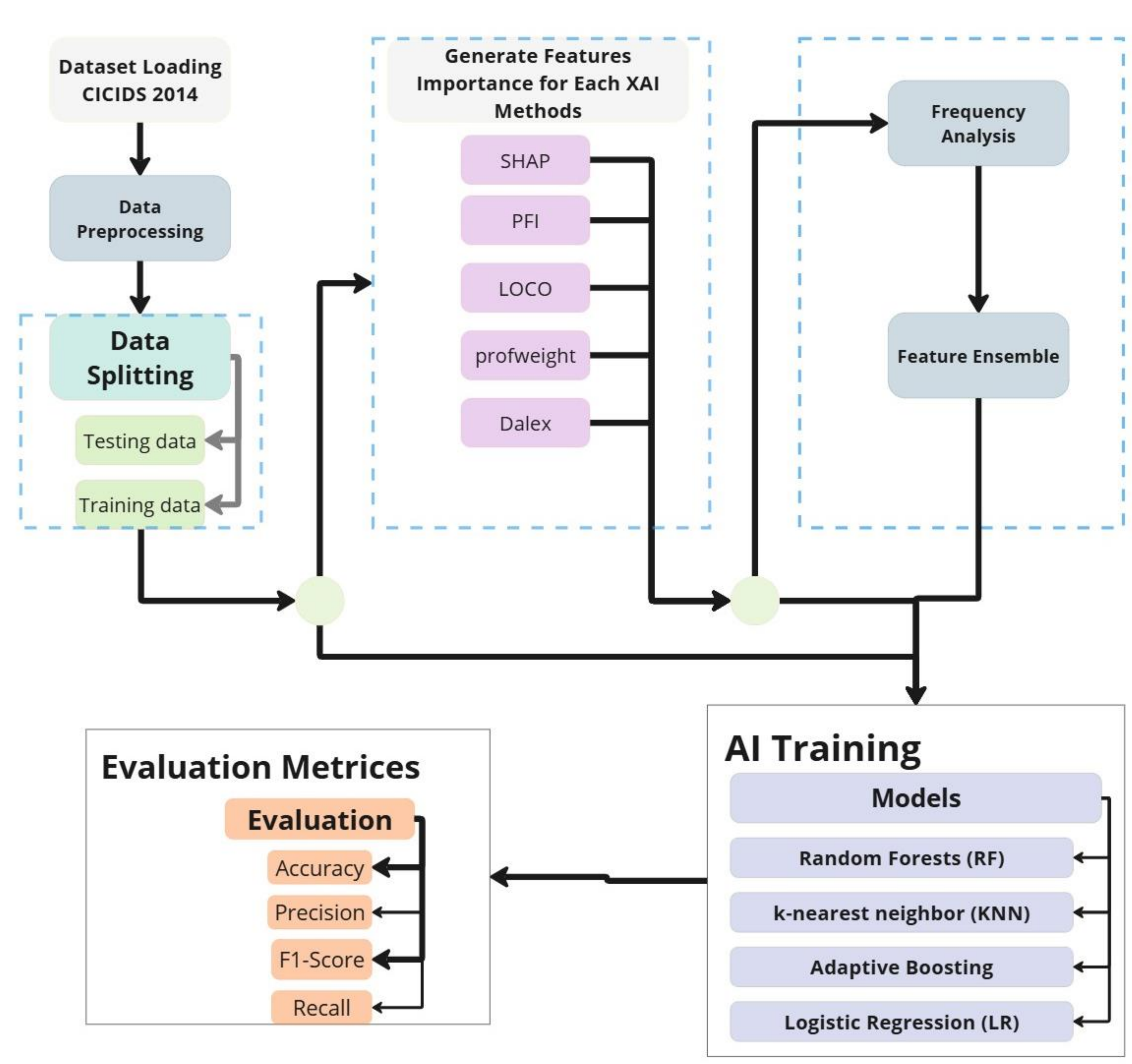
Motivation

- Traditional **Intrusion Detection Systems (IDS)** struggle with high-dimensional data, leading to inefficiencies and reduced detection accuracy.
- **Explainable AI (XAI) methods** help address this issue by providing transparency, enabling better decision-making in network security.
- An **ensemble-based feature selection approach** enhances accuracy, interpretability, and computational efficiency for IDS.

Our Contribution

- We introduce a **frequency-based ensemble feature selection** framework using multiple **XAI methods** (SHAP, LOCO, PFI, DALEX) to enhance intrusion detection.
- We evaluate our approach on the **CICIDS-2017 dataset** with four machine learning models, demonstrating improved **accuracy, precision, and efficiency**.
- We release our **source codes** of the ensemble feature selection framework **to the community**, serving as a baseline for XAI-based feature ensemble in network intrusion detection and encouraging further development with new models.

Framework



XAI Methods – Feature Selection

Dataset

Dataset	Number of Labels	Number of Features	Number of Samples
CICIDS-2017	7	78	2,775,364

Top 10 Features Selection By XAI Methods and Features Ensemble Selection

Rank	SHAP	PROFWEIGHT	LOCO	PFI	DALEX	Feature Ensemble
1	Destination Port	Flow Duration	Init_Win_bytes_forward	Bwd Packet Length Std	Fwd IAT Mean	Fwd IAT Total
2	Init_Win_bytes_forward	Fwd IAT Total	Destination Port	Destination Port	Min Packet Length	Fwd IAT Min
3	Init_Win_bytes_backward	Bwd IAT Total	Flow IAT Min	Init_Win_bytes_forward	Bwd IAT Std	Flow IAT Max
4	min_seg_size_forward	Flow IAT Max	Fwd IAT Min	Total Length of Fwd Packets	Fwd IAT Total	Fwd IAT Mean
5	Flow IAT Min	Fwd IAT Max	Init_Win_bytes_backward	PSH Flag Count	Flow IAT Std	Flow IAT Min
6	Fwd IAT Min	Idle Max	Bwd IAT Std	Bwd Packet Length Min	Bwd Packet Length Max	Fwd Packets
7	Bwd Packets	Idle Mean	Fwd IAT Std	Total Length of Bwd Packets	Max Packet Length	Destination Port
8	Packet Length Mean	Idle Min	Fwd IAT Total	Bwd Packets/s	Active Min	Bwd IAT Std
9	Fwd Packet Length Max	Bwd IAT Max	Flow IAT Std	Average Packet Size	Average Packet Size	Flow Duration
10	Flow Duration	Fwd IAT Std	Fwd Packet Length Max	Packet Length Mean	Fwd Packets	Init_Win_bytes_forward

Results

Configuration	Accuracy	Precision	Recall	F1 Score	Train Time (s)	Predict Time (s)	Total Time (s)
All Features							
All Features	**0.8919**	0.8301	0.8919	0.8480	428.25	4.09	432.34
All Ensemble Features	**0.8919**	0.8301	0.8919	0.8480	429.37	4.10	433.47
K = 10 Features							
PFI	**0.9050**	0.9013	0.9050	0.8906	86.81	2.25	**89.05**
SHAP	0.8839	0.8399	0.8839	0.8556	96.30	2.25	98.55
Ensemble Top 10 features	0.8845	0.8080	0.8845	0.8374	114.00	2.23	116.23
Dalex	0.8861	0.8138	0.8861	0.8422	106.42	2.25	108.67
LOCO	0.8803	0.8096	0.8803	0.8378	94.75	2.24	96.99
ProfWeight	0.8792	0.7905	0.8792	0.8324	102.70	2.24	104.95
K = 5 Features							
PFI	**0.9323**	0.8865	0.9323	0.9083	44.93	2.09	**47.02**
Dalex	0.8816	0.7927	0.8816	0.8344	66.71	2.11	68.83
Ensemble Top 5 Features	0.8799	0.8043	0.8799	0.8335	66.27	2.09	68.36
ProfWeight	0.8773	0.7887	0.8773	0.8304	75.24	2.10	77.34
LOCO	0.8756	0.8237	0.8756	0.8393	49.53	2.08	51.61
SHAP	0.8565	0.7944	0.8565	0.8048	45.98	2.13	48.11

The performance metrics of Adaboost classifier ranked by accuracy

Configuration	Accuracy	Precision	Recall	F1 Score	Train Time (s)	Predict Time (s)	Total Time (s)
All Features							
All Features	**0.9917**	0.9918	0.9917	0.9917	1.14	499.44	500.58
All Ensemble Features	**0.9917**	0.9918	0.9917	0.9917	1.13	498.99	500.12
K = 10 Features							
LOCO	**0.9944**	0.9944	0.9944	0.9944	5.57	120.43	126.00
Dalex	0.9917	0.9915	0.9917	0.9916	5.33	43.23	**48.55**
PFI	0.9886	0.9886	0.9886	0.9883	5.65	58.23	63.88
Ensemble Top 10 Features	0.9883	0.9884	0.9883	0.9883	5.60	47.25	52.86
SHAP	0.9761	0.9763	0.9761	0.9761	5.11	127.27	132.38
ProfWeight	0.9499	0.9426	0.9499	0.9457	5.57	60.56	66.13
K = 5 Features							
PFI	**0.9941**	0.9942	0.9941	0.9936	3.28	181.96	185.24
LOCO	0.9922	0.9926	0.9922	0.9923	3.32	404.18	407.49
SHAP	0.9862	0.9857	0.9862	0.9859	2.92	410.78	413.70
Dalex	0.9722	0.9719	0.9722	0.9693	3.06	282.28	285.34
ProfWeight	0.9456	0.9382	0.9456	0.9413	3.35	37.06	**40.41**
Ensemble Top 5 Features	0.9448	0.9392	0.9448	0.9416	3.37	40.95	44.32

The performance metrics of KNN classifier ranked by accuracy

Configuration	Accuracy	Precision	Recall	F1 Score	Train Time (s)	Predict Time (s)	Total Time (s)
All Features							
All Ensemble Features	**0.8843**	0.8557	0.8843	0.8675	88.93	0.12	89.06
All Features	0.8807	0.8552	0.8807	0.8651	88.14	0.12	**88.27**
K = 10 Features							
PFI	**0.8948**	0.8527	0.8948	0.8673	127.95	0.05	128.00
ProfWeight	0.8706	0.8316	0.8706	0.8348	72.86	0.05	**72.91**
Ensemble Top 10 Features	0.8267	0.8173	0.8267	0.7960	75.36	0.05	75.41
Dalex	0.6970	0.7892	0.6970	0.7282	65.25	0.05	65.30
SHAP	0.7200	0.6732	0.7200	0.6958	69.05	0.05	69.11
LOCO	0.7136	0.7729	0.7136	0.7357	70.20	0.05	70.26
K = 5 Features							
PFI	**0.9185**	0.8770	0.9185	0.8945	72.33	0.05	72.38
SHAP	0.8312	0.6943	0.8312	0.7557	70.97	0.04	71.01
LOCO	0.8191	0.6890	0.8191	0.7484	96.01	0.04	96.05
ProfWeight	0.7046	0.7623	0.7046	0.7167	58.72	0.04	58.76
Ensemble Top 5 Features	0.5662	0.7698	0.5662	0.6208	60.82	0.05	60.87
Dalex	0.5534	0.7717	0.5534	0.6159	55.84	0.05	**55.88**

The performance metrics of Logistic Regression classifier ranked by accuracy

Configuration	Accuracy	Precision	Recall	F1 Score	Train Time (s)	Predict Time (s)	Total Time (s)
All Features							
All Features	**0.9982**	0.9982	0.9982	0.9982	822.80	7.23	830.03
All Ensemble Features	**0.9982**	0.9981	0.9982	0.9982	838.53	7.16	845.69
K = 10 Features							
LOCO	**0.9984**	0.9983	0.9984	0.9984	288.73	3.45	**292.18**
SHAP	0.9983	0.9983	0.9983	0.9983	295.81	3.41	299.23
Dalex	0.9968	0.9967	0.9968	0.9967	433.30	4.39	437.69
PFI	0.9965	0.9966	0.9965	0.9964	311.57	3.56	315.13
Ensemble Top 10 Features	0.9972	0.9972	0.9972	0.9972	385.69	4.02	389.71
ProfWeight	0.9581	0.9475	0.9581	0.9486	462.80	6.00	468.81
K = 5 Features							
PFI	**0.9950**	0.9950	0.9950	0.9944	138.31	3.27	**141.58**
LOCO	0.9947	0.9946	0.9947	0.9946	158.64	3.56	162.20
SHAP	0.9887	0.9882	0.9887	0.9883	136.12	3.67	**139.78**
Ensemble Top 5 Features	0.9541	0.9435	0.9541	0.9446	360.54	6.32	366.86
Dalex	0.9743	0.9741	0.9743	0.9714	295.24	4.72	299.96
ProfWeight	0.9536	0.9428	0.9536	0.9440	412.70	6.05	418.75

The performance metrics of Random Forest classifier ranked by accuracy

Key Takeaways

- Development of a **novel frequency-based ensemble feature selection framework**, integrating multiple **XAI techniques** to improve feature ranking in network IDS.
- The study demonstrates that combining **multiple XAI-based rankings** enhances model performance, providing a more reliable approach to feature selection for IDS

Acknowledgement

This work was supported by AnalytixIN, Enhanced Mentoring Program with Opportunities for Ways to Excel in Research (EMPOWER); By the 1st Year Research Immersion Program (1RIP) Grants from the Office of the Vice Chancellor for Research at Purdue University Indianapolis; and in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute.

PURDUE UNIVERSITY

CERIAS