AI/ML FOR NETWORKING

AI NETWORK TRAFFIC ANALYSIS

PRESENTATION

TEAM:

Jyoshika Balam

VU22CSEN0100432

Sushanth K

VU22CSEN0100016

Problem Summary:

Modern networks face increasing challenges in monitoring and securing traffic due to the exponential growth of data, encrypted communication, and sophisticated cyber threats. This presentation explores how AI and machine learning can revolutionize network security and operations.

Modern networks = Huge Data + Encryption + Advanced Threats → Old methods are not enough.

- Traditional rule-based security and DPI are failing.
- Manual classification is slow and error-prone.

➞ Need AI-based real-time analysis to:

- Detect patterns.
- Spot anomalies.
- Secure encrypted traffic.

Deliverables:

1.Traffic Classification Model

- AI model that identifies types of traffic (e.g., WhatsApp, YouTube, VPNs) from packet metadata or flow patterns.

2. Threat Detection & Anomaly Framework

AI system that detects:

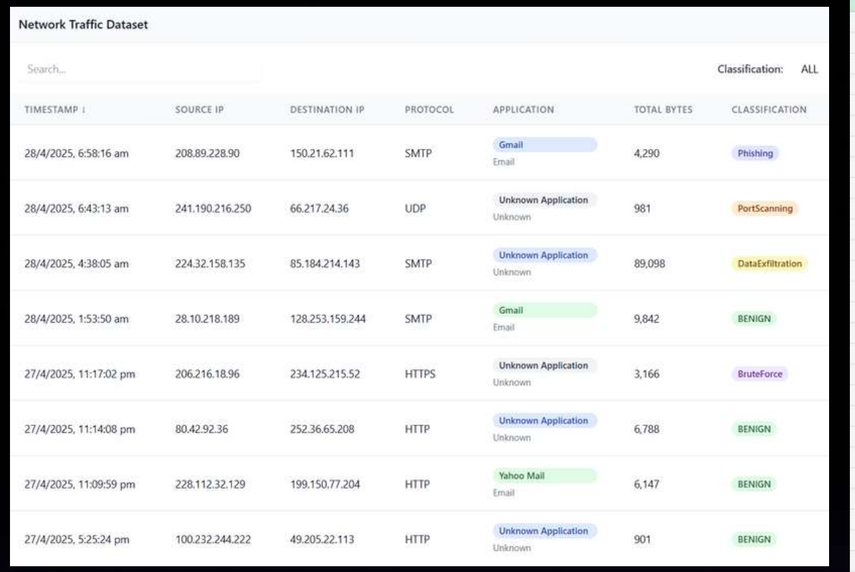
- Malware.
- DDoS attacks.
- Suspicious behaviors.
- Unknown anomalies.

Model Information

ARCHITECTURE	Ensemble Model: Random Forest + XGBoost
FEATURES USED	32 features
DATA SPLIT	70% Train / 15% Validation / 15% Test
OBJECTIVE	Classify network traffic and detect threats/anomalies
ALGORITHM	Multi-class Classification
EVALUATION METRICS	Accuracy, Precision, Recall, F1 Score, Confusion Matrix
TOP FEATURES	Flow Duration | Total Bytes | Packet Inter-arrival Time (Mean) | Flow Bytes per Second | TCP Flags

DATASET

- [NETWORK_TRAFFIC_DATA_2025-04-28.CSV]

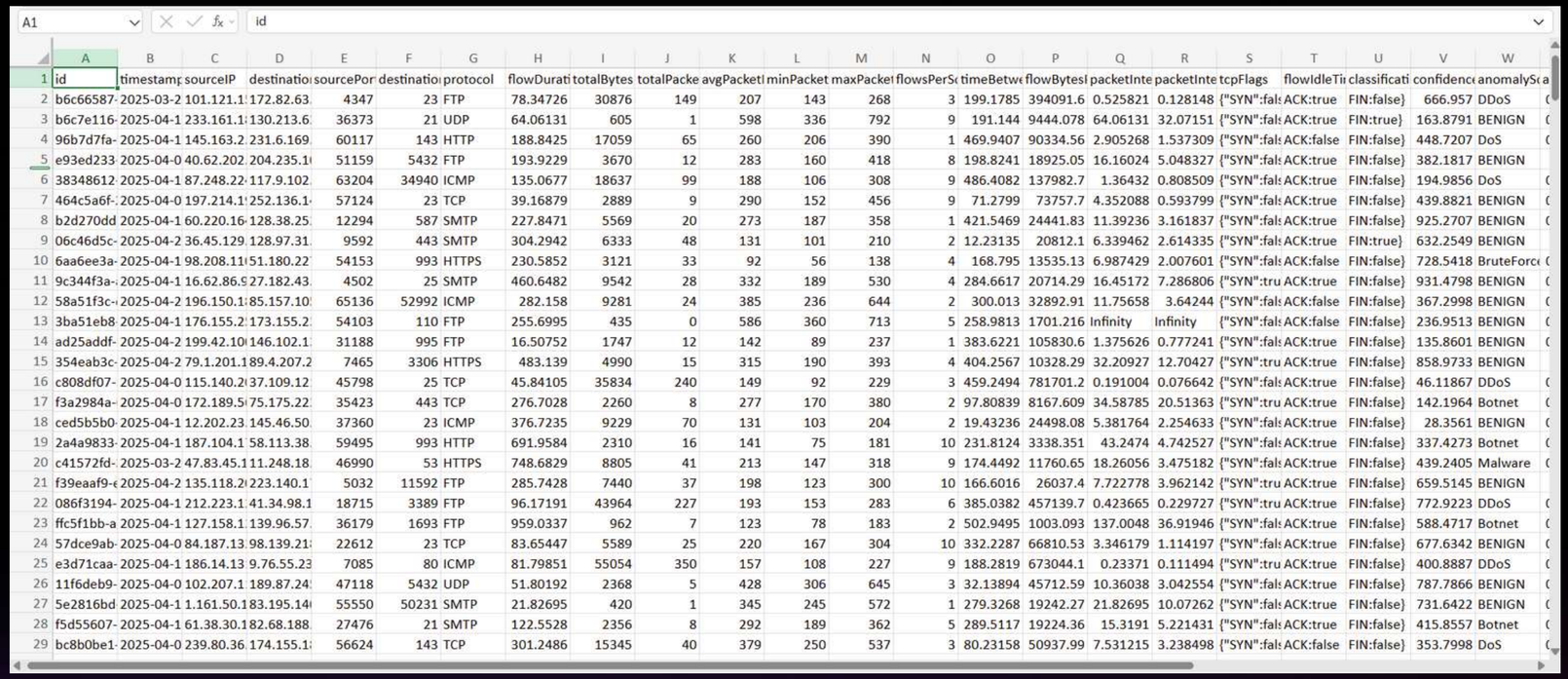


Network Traffic Dataset

Search...

Classification: ALL

TIMESTAMP	SOURCE IP	DESTINATION IP	PROTOCOL	APPLICATION	TOTAL BYTES	CLASSIFICATION
28/4/2025, 6:58:16 am	208.89.228.90	150.21.62.111	SMTP	Gmail Email	4,290	Phishing
28/4/2025, 6:43:13 am	241.190.216.250	66.217.24.36	UDP	Unknown Application Unknown	981	PortScanning
28/4/2025, 4:38:05 am	224.32.158.135	85.184.214.143	SMTP	Unknown Application Unknown	89,098	DataExfiltration
28/4/2025, 1:53:50 am	28.10.218.189	128.253.159.244	SMTP	Gmail Email	9,842	BENIGN
27/4/2025, 11:17:02 pm	206.216.18.96	234.125.215.52	HTTPS	Unknown Application Unknown	3,166	BruteForce
27/4/2025, 11:14:08 pm	80.42.92.36	252.36.65.208	HTTP	Unknown Application Unknown	6,788	BENIGN
27/4/2025, 11:09:59 pm	228.112.32.129	199.150.77.204	HTTP	Yahoo Mail Email	6,147	BENIGN
27/4/2025, 5:25:24 pm	100.232.244.222	49.205.22.113	HTTP	Unknown Application Unknown	901	BENIGN



A1 | id

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	id	timestamp	sourceIP	destinatio	sourcePor	destinatio	protocol	flowDurati	totalBytes	totalPacke	avgPacket	minPacket	maxPacket	flowsPerSe	timeBetwe	flowBytesI	packetInte	packetInte	tcpFlags	flowIdleTi	classificati	confidenc	anomalySc
2	b6c66587-	2025-03-2	101.121.1	172.82.63	4347	23	FTP	78.34726	30876	149	207	143	268	3	199.1785	394091.6	0.525821	0.128148	{"SYN":fals	ACK:true	FIN:false}	666.957	DDoS
3	b6c7e116-	2025-04-1	233.161.1	130.213.6	36373	21	UDP	64.06131	605	1	598	336	792	9	191.144	9444.078	64.06131	32.07151	{"SYN":fals	ACK:true	FIN:true}	163.8791	BENIGN
4	96b7d7fa-	2025-04-1	145.163.2	231.6.169	60117	143	HTTP	188.8425	17059	65	260	206	390	1	469.9407	90334.56	2.905268	1.537309	{"SYN":fals	ACK:false	FIN:false}	448.7207	DoS
5	e93ed233	2025-04-0	40.62.202	204.235.1	51159	5432	FTP	193.9229	3670	12	283	160	418	8	198.8241	18925.05	16.16024	5.048327	{"SYN":fals	ACK:true	FIN:false}	382.1817	BENIGN
6	38348612	2025-04-1	87.248.22	117.9.102	63204	34940	ICMP	135.0677	18637	99	188	106	308	9	486.4082	137982.7	1.36432	0.808509	{"SYN":fals	ACK:true	FIN:false}	194.9856	DoS
7	464c5a6f-	2025-04-0	197.214.1	252.136.1	57124	23	TCP	39.16879	2889	9	290	152	456	9	71.2799	73757.7	4.352088	0.593799	{"SYN":fals	ACK:true	FIN:false}	439.8821	BENIGN
8	b2d270dd	2025-04-1	60.220.16	128.38.25	12294	587	SMTP	227.8471	5569	20	273	187	358	1	421.5469	24441.83	11.39236	3.161837	{"SYN":fals	ACK:true	FIN:false}	925.2707	BENIGN
9	06c46d5c-	2025-04-2	36.45.129	128.97.31	9592	443	SMTP	304.2942	6333	48	131	101	210	2	12.23135	20812.1	6.339462	2.614335	{"SYN":fals	ACK:true	FIN:true}	632.2549	BENIGN
10	6aa6ee3a-	2025-04-1	98.208.11	51.180.22	54153	993	HTTPS	230.5852	3121	33	92	56	138	4	168.795	13535.13	6.987429	2.007601	{"SYN":fals	ACK:false	FIN:false}	728.5418	BruteForce
11	9c344f3a-	2025-04-1	16.62.86.9	27.182.43	4502	25	SMTP	460.6482	9542	28	332	189	530	4	284.6617	20714.29	16.45172	7.286806	{"SYN":tru	ACK:true	FIN:false}	931.4798	BENIGN
12	58a51f3c-	2025-04-2	196.150.1	85.157.10	65136	52992	ICMP	282.158	9281	24	385	236	644	2	300.013	32892.91	11.75658	3.64244	{"SYN":fals	ACK:false	FIN:false}	367.2998	BENIGN
13	3ba51eb8	2025-04-1	176.155.2	173.155.2	54103	110	FTP	255.6995	435	0	586	360	713	5	258.9813	1701.216	Infinity	Infinity	{"SYN":fals	ACK:false	FIN:false}	236.9513	BENIGN
14	ad25addf-	2025-04-2	199.42.10	146.102.1	31188	995	FTP	16.50752	1747	12	142	89	237	1	383.6221	105830.6	1.375626	0.777241	{"SYN":fals	ACK:true	FIN:false}	135.8601	BENIGN
15	354eab3c-	2025-04-2	79.1.201.1	89.4.207.2	7465	3306	HTTPS	483.139	4990	15	315	190	393	4	404.2567	10328.29	32.20927	12.70427	{"SYN":tru	ACK:true	FIN:false}	858.9733	BENIGN
16	c808df07-	2025-04-0	115.140.2	37.109.12	45798	25	TCP	45.84105	35834	240	149	92	229	3	459.2494	781701.2	0.191004	0.076642	{"SYN":fals	ACK:true	FIN:false}	46.11867	DDoS
17	f3a2984a-	2025-04-0	172.189.5	75.175.22	35423	443	TCP	276.7028	2260	8	277	170	380	2	97.80839	8167.609	34.58785	20.51363	{"SYN":tru	ACK:true	FIN:false}	142.1964	Botnet
18	ced5b5b0	2025-04-1	12.202.23	145.46.50	37360	23	ICMP	376.7235	9229	70	131	103	204	2	19.43236	24498.08	5.381764	2.254633	{"SYN":fals	ACK:true	FIN:false}	28.3561	BENIGN
19	2a4a9833-	2025-04-1	187.104.1	58.113.38	59495	993	HTTP	691.9584	2310	16	141	75	181	10	231.8124	3338.351	43.2474	4.742527	{"SYN":fals	ACK:true	FIN:false}	337.4273	Botnet
20	c41572fd-	2025-03-2	47.83.45.1	111.248.18	46990	53	HTTPS	748.6829	8805	41	213	147	318	9	174.4492	11760.65	18.26056	3.475182	{"SYN":fals	ACK:true	FIN:false}	439.2405	Malware
21	f39eaaf9-e	2025-04-2	135.118.2	223.140.1	5032	11592	FTP	285.7428	7440	37	198	123	300	10	166.6016	26037.4	7.722778	3.962142	{"SYN":tru	ACK:true	FIN:false}	659.5145	BENIGN
22	086f3194-	2025-04-1	212.223.1	41.34.98.1	18715	3389	FTP	96.17191	43964	227	193	153	283	6	385.0382	457139.7	0.423665	0.229727	{"SYN":fals	ACK:true	FIN:false}	772.9223	DDoS
23	ffc5f1bb-a	2025-04-1	127.158.1	139.96.57	36179	1693	FTP	959.0337	962	7	123	78	183	2	502.9495	1003.093	137.0048	36.91946	{"SYN":fals	ACK:true	FIN:false}	588.4717	Botnet
24	57dce9ab-	2025-04-0	84.187.13	98.139.21	22612	23	TCP	83.65447	5589	25	220	167	304	10	332.2287	66810.53	3.346179	1.114197	{"SYN":fals	ACK:true	FIN:false}	677.6342	BENIGN
25	e3d71caa-	2025-04-1	186.14.13	9.76.55.23	7085	80	ICMP	81.79851	55054	350	157	108	227	9	188.2819	673044.1	0.23371	0.111494	{"SYN":tru	ACK:true	FIN:false}	400.8887	DDoS
26	11f6deb9-	2025-04-0	102.207.1	189.87.24	47118	5432	UDP	51.80192	2368	5	428	306	645	3	32.13894	45712.59	10.36038	3.042554	{"SYN":fals	ACK:true	FIN:false}	787.7866	BENIGN
27	5e2816bd	2025-04-1	1.161.50.1	83.195.14	55550	50231	SMTP	21.82695	420	1	345	245	572	1	279.3268	19242.27	21.82695	10.07262	{"SYN":fals	ACK:true	FIN:false}	731.6422	BENIGN
28	f5d55607-	2025-04-1	61.38.30.1	82.68.188	27476	21	SMTP	122.5528	2356	8	292	189	362	5	289.5117	19224.36	15.3191	5.221431	{"SYN":fals	ACK:true	FIN:false}	415.8557	Botnet
29	bc8b0be1-	2025-04-0	239.80.36	174.155.1	56624	143	TCP	301.2486	15345	40	379	250	537	3	80.23158	50937.99	7.531215	3.238498	{"SYN":fals	ACK:false	FIN:false}	353.7998	DoS

- **Data set is attached to the MAIL**

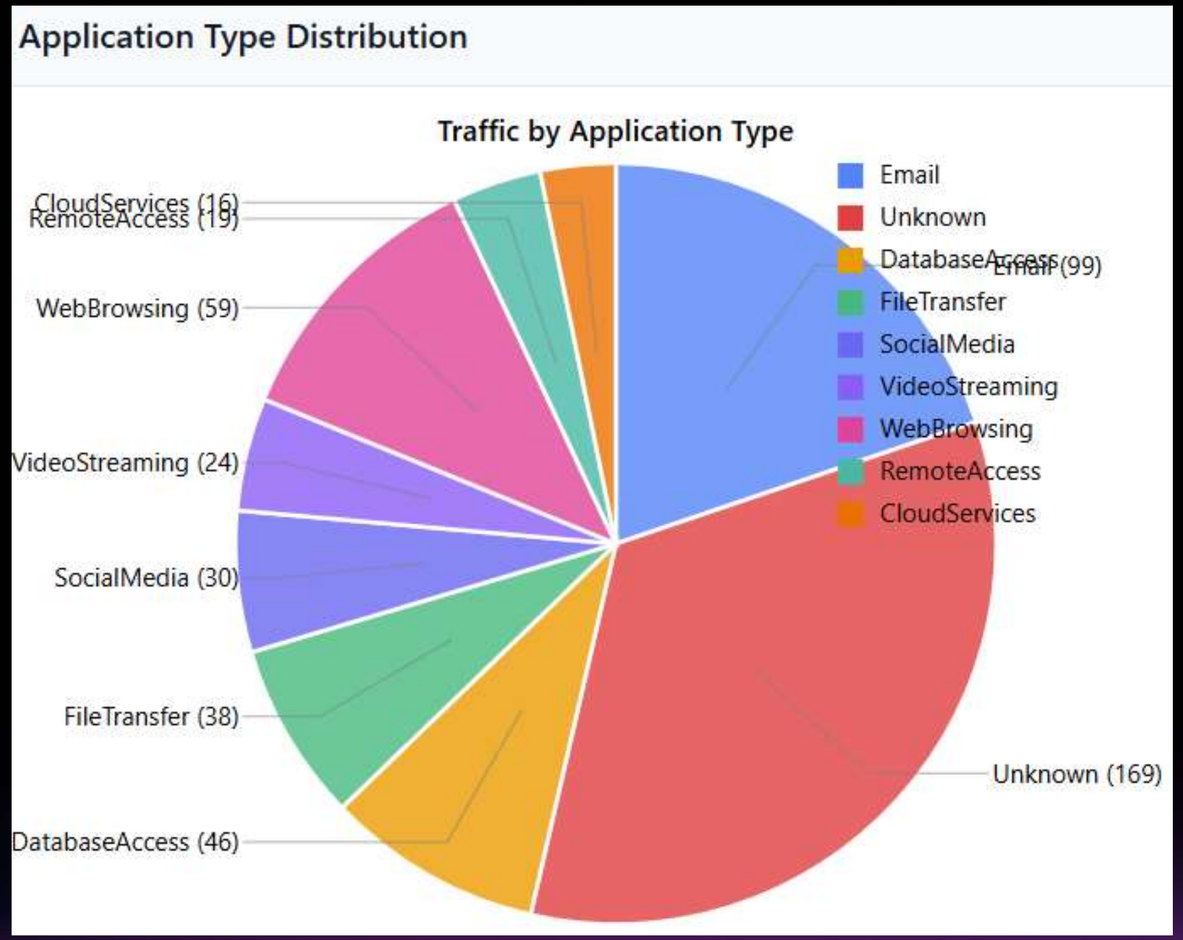
Traffic Classification Model

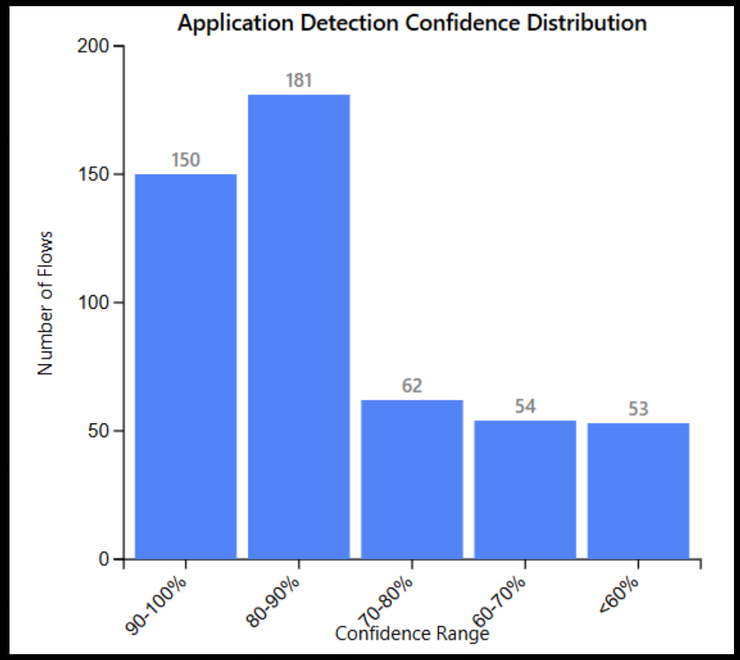
Detected applications include:

1. Email
2. Web Browsing
3. Social Media
4. File Transfer
5. Database Access
6. Video Streaming
7. Remote Access
8. Cloud Services
9. Unknown Applications

- Web browsing, email, and unknown traffic are among the most commonly observed application types.
- Remote access and cloud services were identified less frequently.

Classification of applications:





Application Detection Confidence:

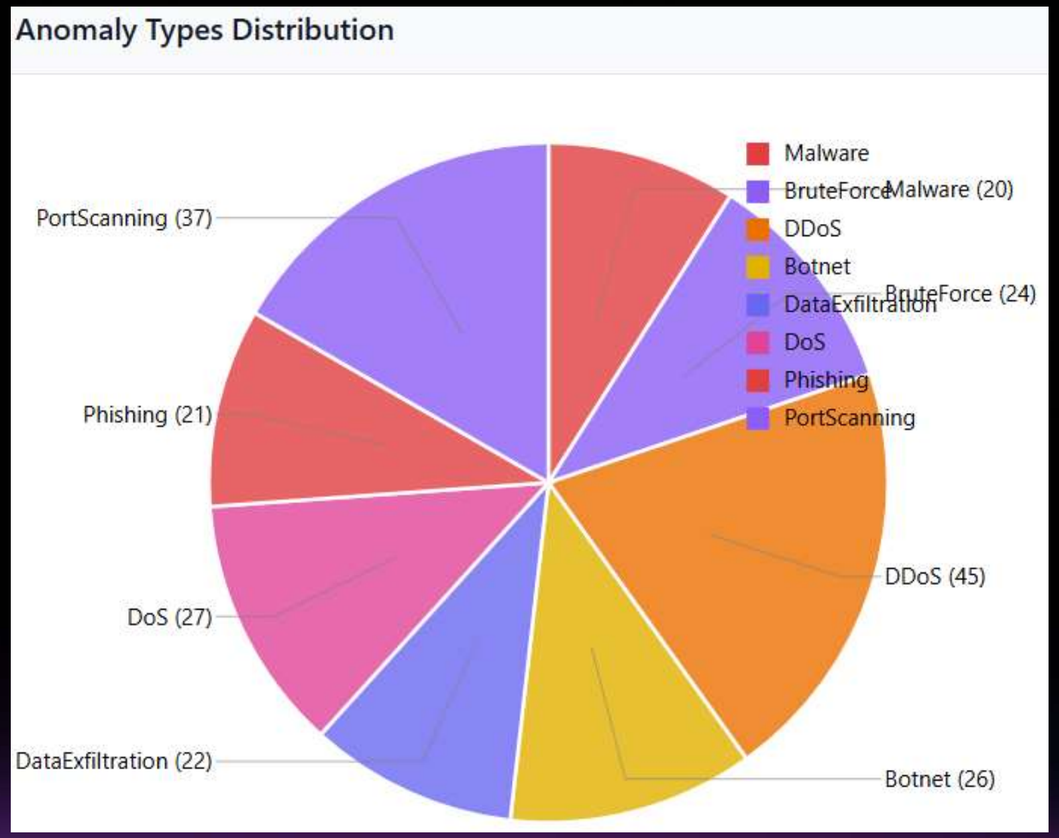
- Majority of application detections achieved high confidence levels (above 80%).
- A smaller proportion falls into lower confidence ranges (below 70%).
- Overall, application identification is considered reliable and robust.

Threat Detection & Anomaly Identification Framework

Detected anomaly categories include:

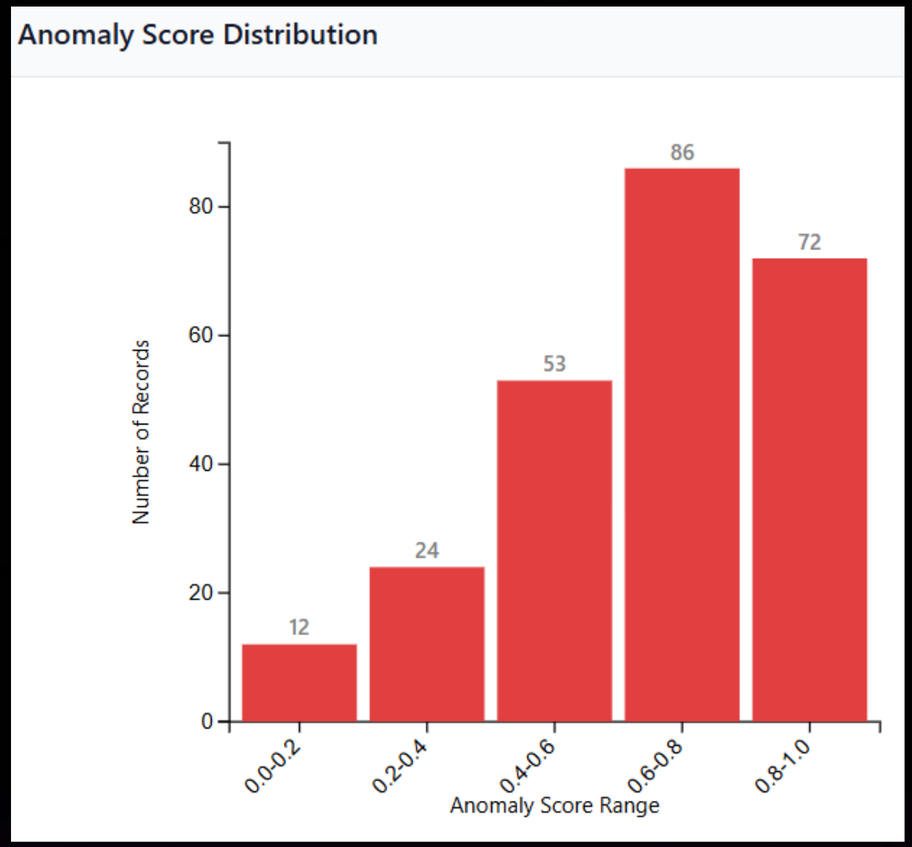
1. Malware
2. Brute Force Attacks
3. Distributed Denial of Service (DDoS)
4. Botnet Activity
5. Data Exfiltration Attempts
6. Denial of Service (DoS)
7. Phishing Attempts
8. Port Scanning Activities

All major types of network anomalies were effectively identified and classified.



Anomaly Score Distribution:

- Records were distributed across various anomaly score ranges.
- The majority of anomalies have moderate to high anomaly scores (between 0.4 and 1.0).
- Very few records fall into low anomaly score ranges (below 0.4).
- This distribution reflects the system's strong capability to distinguish between benign and truly malicious activities.

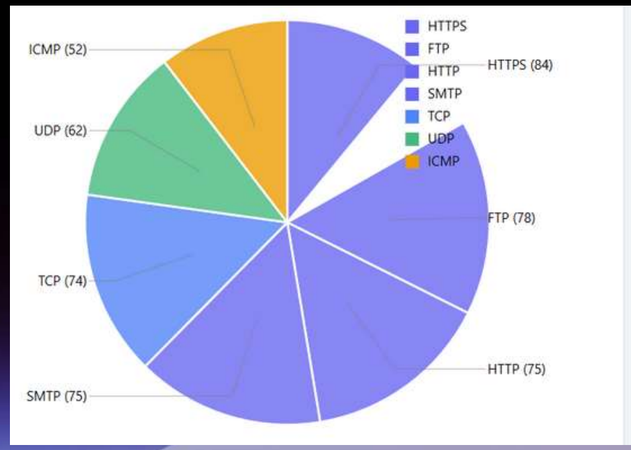


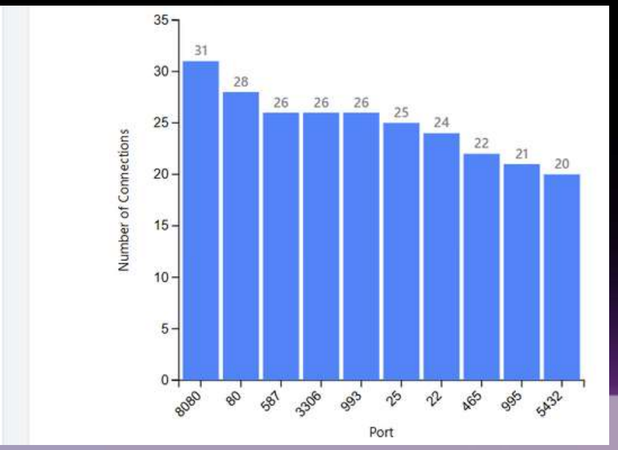
Traffic by Protocol

Network traffic was observed across multiple protocols, including:

1. HTTPS
2. FTP
3. HTTP
4. SMTP
5. TCP
6. UDP
7. ICMP

This diversity in protocols highlights a wide variety of network activities, both secure and unsecure.





Traffic by Port (Top 10)

Top ports identified in the network flows include:

1. 8080 (Alternative HTTP port)
2. 80 (Standard HTTP)
3. 587 (SMTP Secure Email Submission)
4. 3306 (MySQL Database)
5. 993 (IMAP Secure Email Retrieval)
6. 25 (SMTP Email Transfer)
7. 22 (SSH Remote Login)
8. 465 (SMTPS Secure Email Submission)
9. 995 (POP3 Secure Email Retrieval)
10. 5432 (PostgreSQL Database)

Traffic across these ports suggests a mix of web services, email communications, database interactions, and secure remote access.

Ensemble Model Architecture

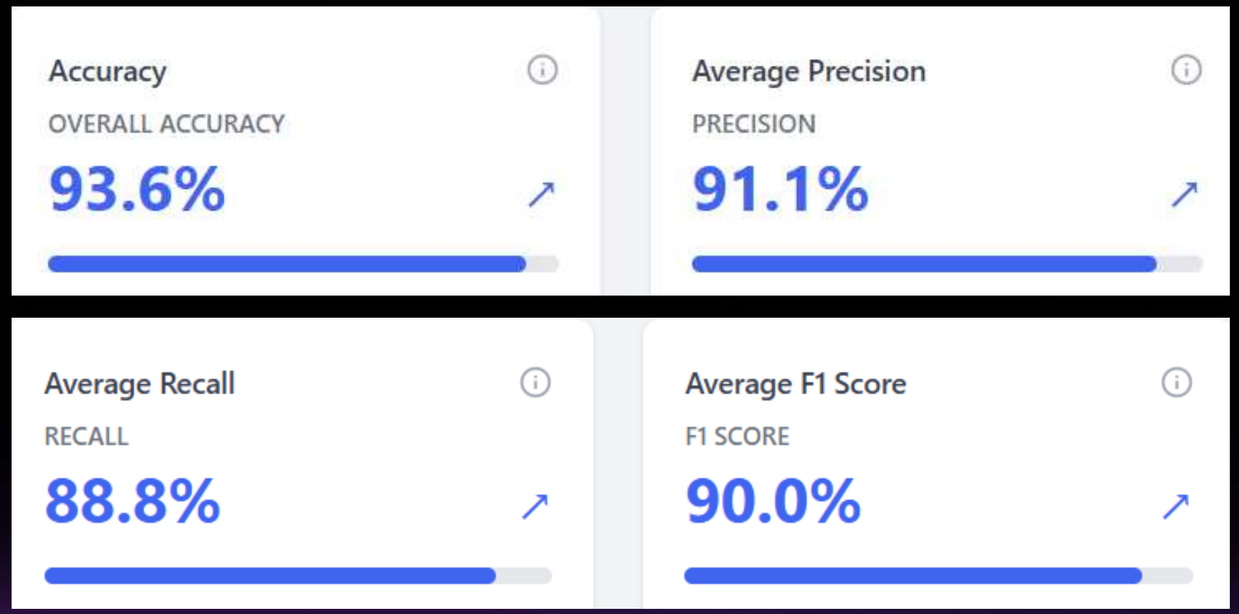
For this project, we used an ensemble model, which combines the strengths of multiple classifiers to improve prediction accuracy and reduce errors. In particular, we used:

- Random Forest (a decision tree-based model) for its strength in handling high-dimensional data.
- XGBoost (a gradient boosting method) to fine-tune the model for better performance.

This combination allows the model to learn from diverse perspectives and make more accurate predictions than individual models alone.

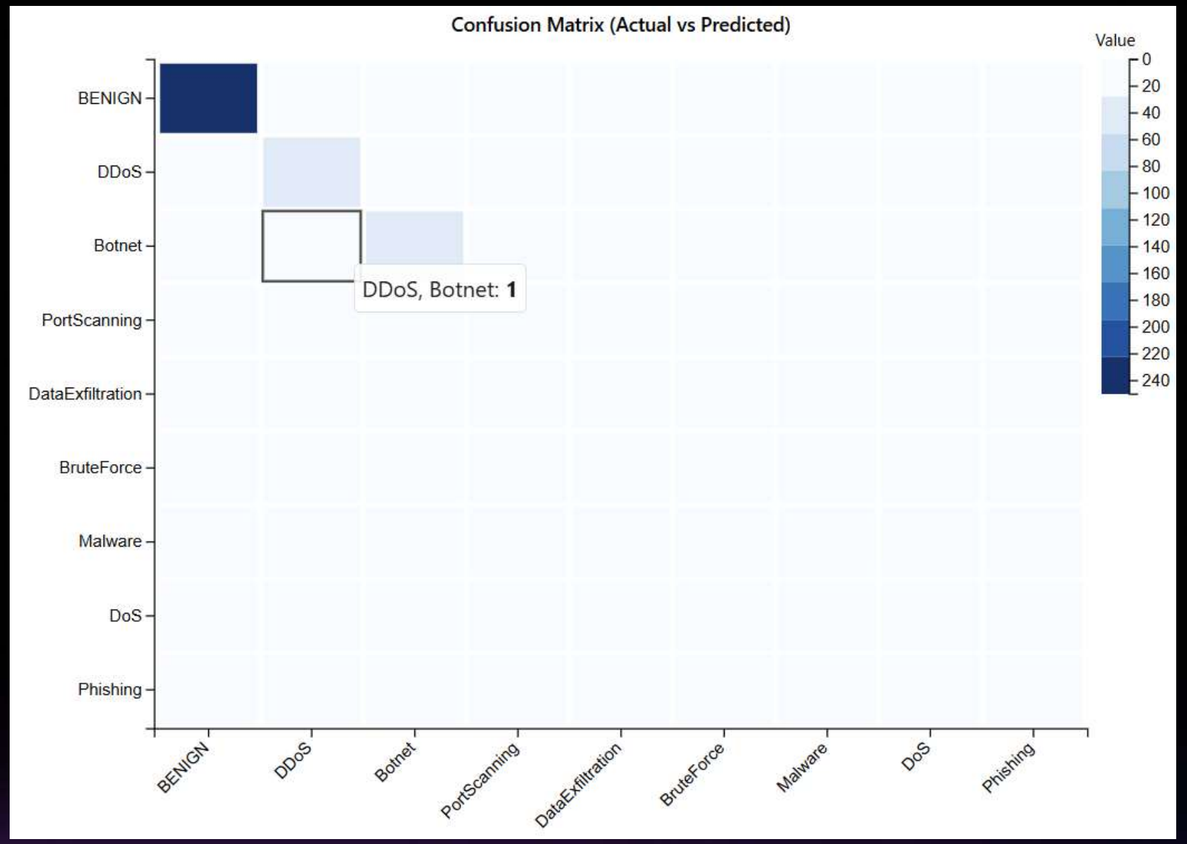
Model Metrics

Accuracy:
93.6%
Precision: 91.1%
Recall: 88.8%
F1 Score: 90.0%



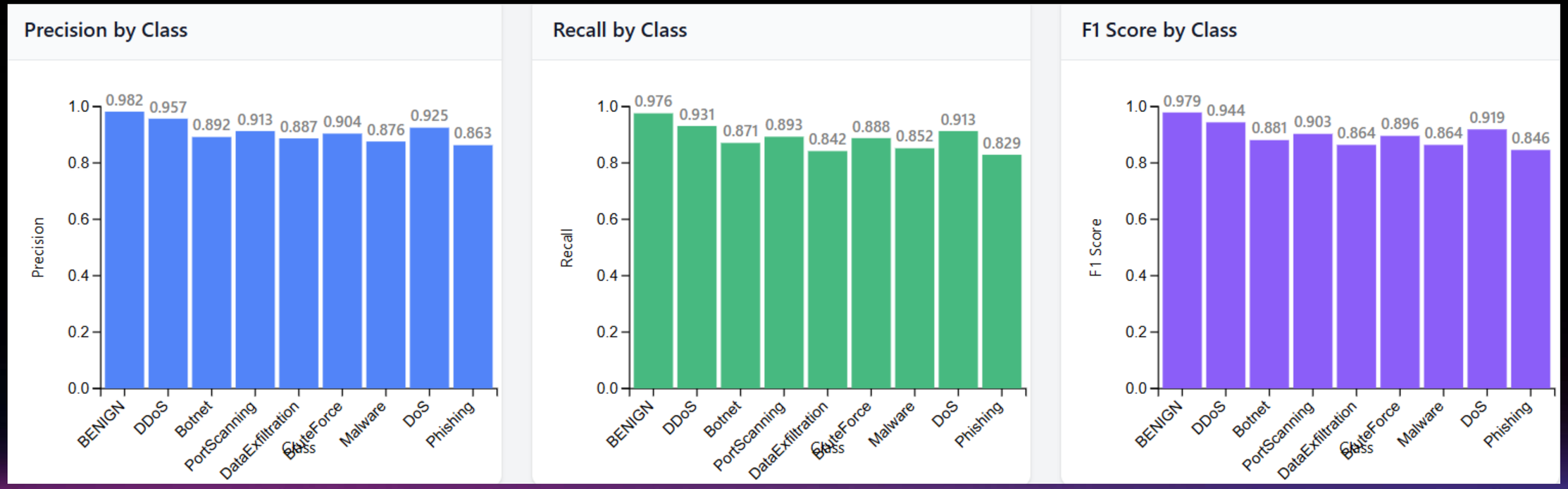
Confusion Matrix

- The confusion matrix shows how well the model classifies traffic across different categories:
- True Positives: The model correctly identified benign traffic and malicious activities (e.g., DDoS, Malware).
- False Positives: Misclassifications of benign traffic as threats.
- False Negatives: Failure to detect some threats.
- The ensemble model shows strong performance, especially in detecting benign traffic and common attacks like DDoS and Malware.



Precision, Recall, and F1 Score by Class:

Precision is particularly high for Benign traffic, DDoS, and Botnet traffic, which indicates that the model rarely misclassifies benign traffic or fails to detect malicious activity.
Recall (True Positives) is high across most categories, with Botnet and Port Scanning showing particularly robust detection.
The F1 Score balances the precision and recall to ensure high reliability across all traffic types



Outcomes:

- Automated Network Traffic Analysis using AI/ML models to detect and classify traffic in real time.
- Improved Threat Detection & Security, identifying anomalies, malware, and encrypted attacks with higher accuracy.
- Reduced False Positives & False Negatives, enhancing the efficiency of network security operations.
- Scalability & Performance Optimization, ensuring AI models can handle high-traffic environments with minimal latency.
- Privacy-Preserving Traffic Analysis, leveraging AI for encrypted traffic analysis without decryption.

THANK YOU!