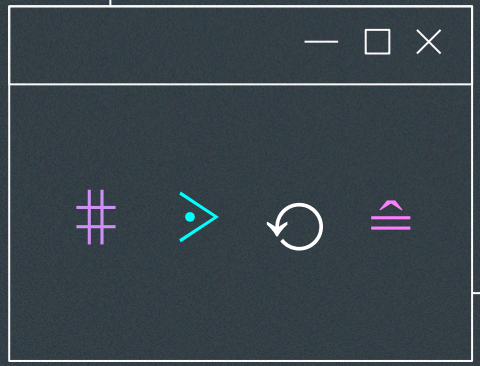
Comparação de Modelos Classificadores na Detecção de Anomalias em Sistemas de Detecção de Intrusão

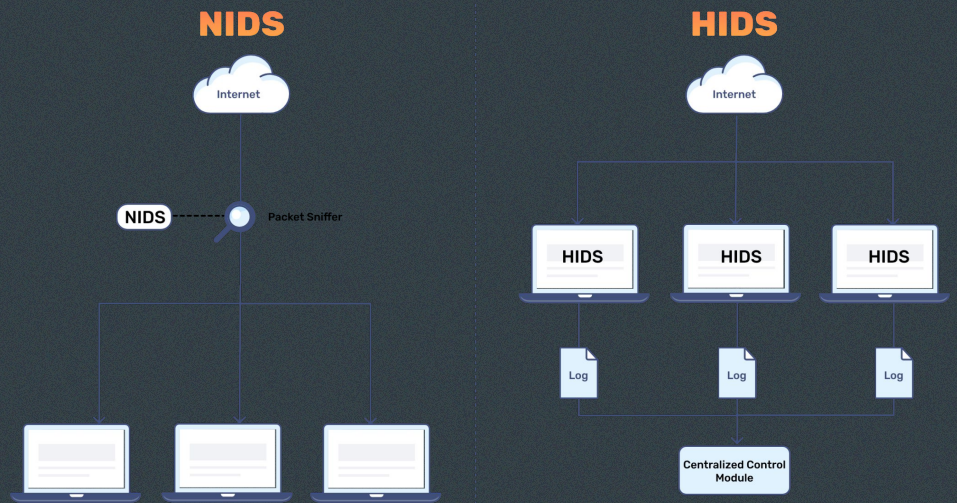


Nome: Lázaro Robert da Silva Cunha

Sumário

Introdução

- O que é um Sistema de Detecção de Intrusão?
- Tipos: host e redes sem fio
- Identificam ataques por assinatura ou por anomalia



Motivação e Objetivo

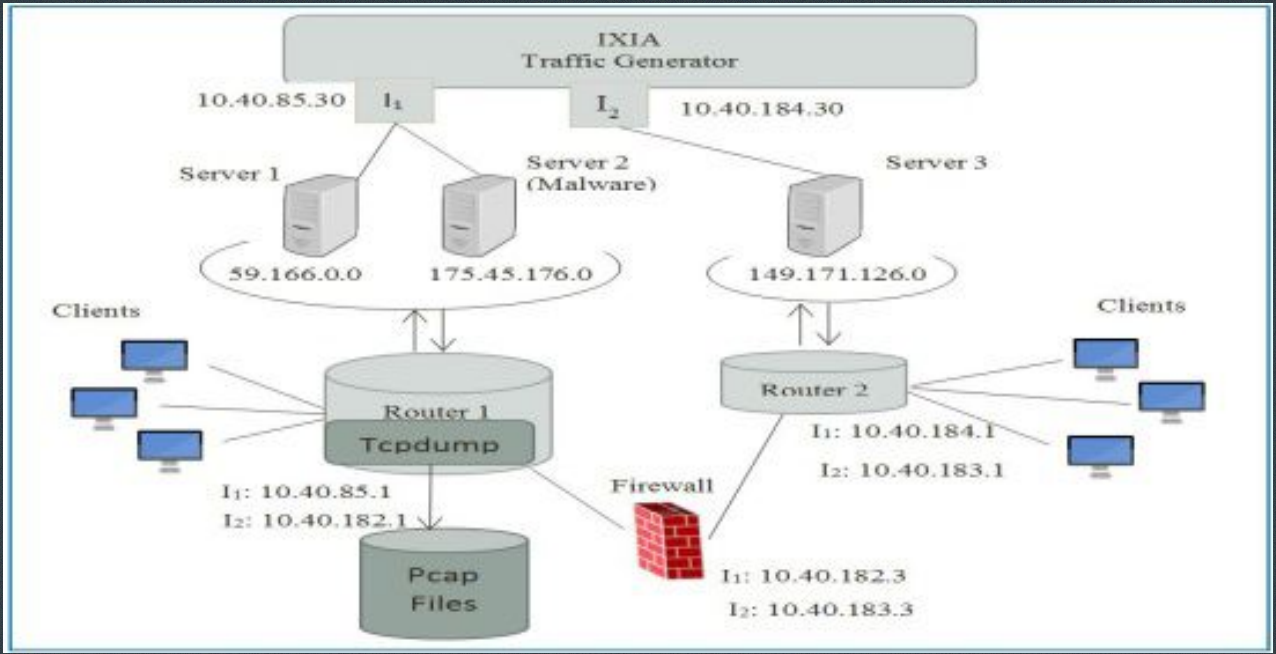
- Por que é interessante?
- Expansão da digitalização
- Aumento das ameaças

- Comparação de modelos de aprendizagem de máquina.
- Encontrar o mais adequado.

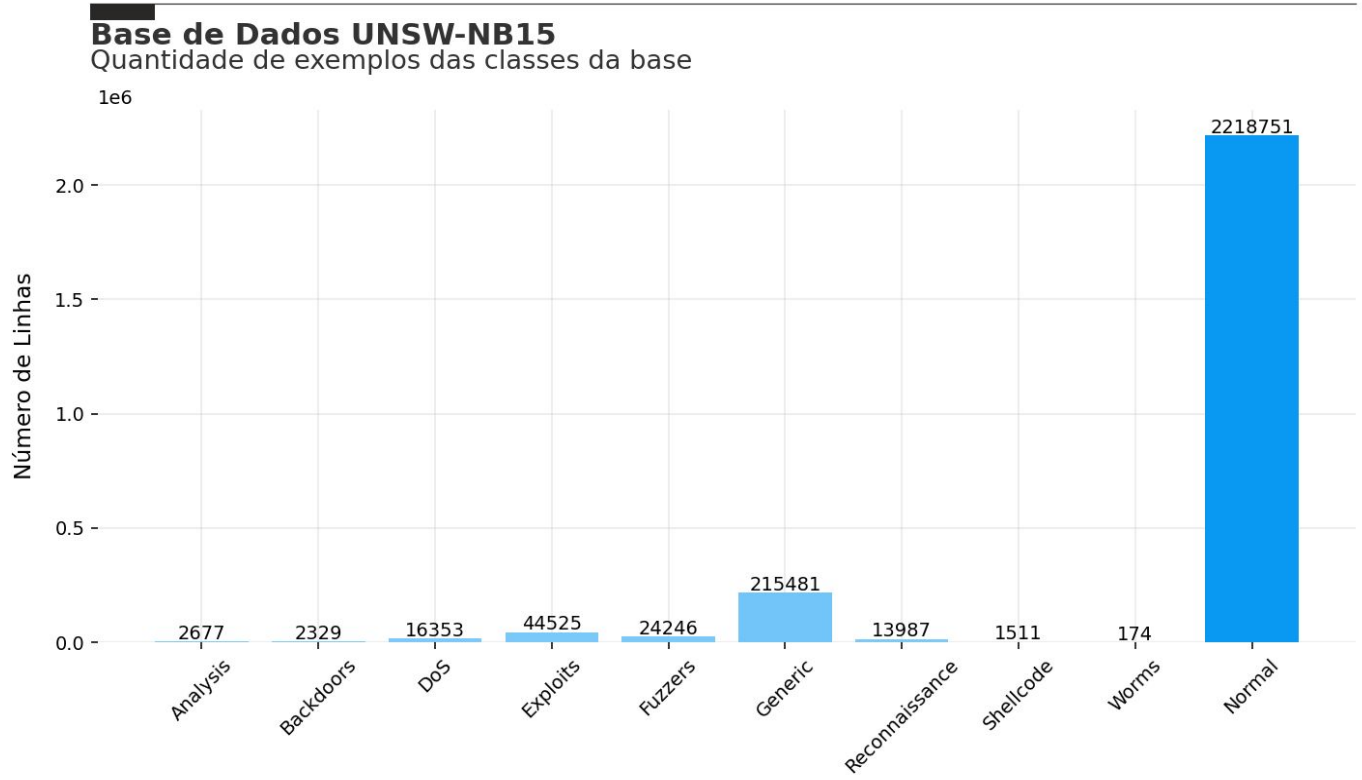
Metodologia

- Ferramentas: Jupyter, Pandas, Numpy, Scikit-Learn e HiClass
- Base de dados: UNSW-NB15
    - 2.5M de exemplos
    - 47 Características

Metodologia



Metodologia



Pré-processamento dos dados

- União do conjunto
- Busca por inconsistências:
  - Dados faltantes
  - Inconsistências de tipos e formatos
- Normalização

Seleção de Características

- GridSearchCV no Random Forest
- Treino do modelo em Cross Val
- Média das importâncias das características

Feature	Random Forest Feature Importance
dstip	0.171892
sttl	0.171047
ct_state_ttl	0.158517
sbytes	0.085213
srcip	0.080752
dmeansz	0.031187
smeansz	0.030895
dbytes	0.025733

Execução dos Modelos

- Modelos executados:
  - Adaptive Boosting
  - Random Forest
  - XGBoost
  - Bagging
  - MLP
  - Classificador Local por Nó, por Nó Pai e por Nível.
- Validação cruzada
- Métricas coletadas:
  - acurácia, precisão, recall, f1-score

Resultados

Bagging

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Normal	0.986528	0.991463	0.999716	0.984851	0.992228
Attack	0.986528	0.991463	0.905347	0.998075	0.949452

Adaptive Boosting

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Normal	0.99008	0.984432	0.996627	0.991997	0.994306
Attack	0.99008	0.984432	0.946629	0.976867	0.961496

MLP

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Normal	0.989932	0.981786	0.995765	0.992697	0.994227
Attack	0.989932	0.981786	0.950908	0.970876	0.960703

Resultados

Random Forest

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Normal	0.992835	0.981564	0.995139	0.996663	0.995900
Attack	0.992835	0.981564	0.976767	0.966465	0.971588

Extreme Gradient Boosting

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Normal	0.992021	0.982725	0.995682	0.995178	0.995430
Attack	0.992021	0.982725	0.966901	0.970272	0.968583

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Analysis	0.873237	0.1	NaN	0.0	0.000000
Backdoors	0.873237	0.1	NaN	0.0	0.000000
DoS	0.873237	0.1	NaN	0.0	0.000000
Exploits	0.873237	0.1	NaN	0.0	0.000000
Fuzzers	0.873237	0.1	NaN	0.0	0.000000
Generic	0.873237	0.1	NaN	0.0	0.000000
Reconnaissance	0.873237	0.1	NaN	0.0	0.000000
Shellcode	0.873237	0.1	NaN	0.0	0.000000
Worms	0.873237	0.1	NaN	0.0	0.000000
Normal	0.873237	0.1	0.873237	1.0	0.932329

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Analysis	0.963769	0.42039	0.049700	0.460048	0.089426
Backdoors	0.963769	0.42039	0.033755	0.036515	0.025713
DoS	0.963769	0.42039	0.247432	0.138318	0.135973
Exploits	0.963769	0.42039	0.643191	0.561701	0.585021
Fuzzers	0.963769	0.42039	0.357247	0.319173	0.321097
Generic	0.963769	0.42039	0.987042	0.974678	0.980808
Reconnaissance	0.963769	0.42039	0.560558	0.650585	0.599229
Shellcode	0.963769	0.42039	0.000000	0.000000	0.000000
Worms	0.963769	0.42039	0.069811	0.074762	0.062739
Normal	0.963769	0.42039	0.991638	0.988117	0.989862

Bagging, Adaptive Boosting, MLP e Random Forest

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Analysis	0.979068	0.510448	0.609618	0.023673	0.044621
Backdoors	0.979068	0.510448	0.644488	0.059450	0.106346
DoS	0.979068	0.510448	0.417983	0.330088	0.273759
Exploits	0.979068	0.510448	0.658738	0.766836	0.687759
Fuzzers	0.979068	0.510448	0.670194	0.437320	0.523099
Generic	0.979068	0.510448	0.993927	0.983560	0.988716
Reconnaissance	0.979068	0.510448	0.819709	0.754799	0.784980
Shellcode	0.979068	0.510448	0.515739	0.508747	0.504477
Worms	0.979068	0.510448	0.360000	0.242540	0.221685
Normal	0.979068	0.510448	0.994035	0.997467	0.995748

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Analysis	0.983547	0.584948	0.871317	0.121515	0.212761
Backdoors	0.983547	0.584948	0.817051	0.097931	0.174462
DoS	0.983547	0.584948	0.309749	0.235063	0.267138
Exploits	0.983547	0.584948	0.631257	0.824145	0.714890
Fuzzers	0.983547	0.584948	0.779630	0.683635	0.728400
Generic	0.983547	0.584948	0.997990	0.987289	0.992611
Reconnaissance	0.983547	0.584948	0.925980	0.780141	0.846814
Shellcode	0.983547	0.584948	0.844426	0.901570	0.871401
Worms	0.983547	0.584948	0.644286	0.219683	0.323480
Normal	0.983547	0.584948	0.997471	0.998512	0.997991

Resultados

Extreme Gradient Boosting

	Acurácia no Teste	Acurácia Balanceada no Teste	Precisão	Recall	F1
Analysis	0.98352	0.610389	0.848095	0.084735	0.153981
Backdoors	0.98352	0.610389	0.893626	0.096329	0.173710
DoS	0.98352	0.610389	0.434234	0.300773	0.355235
Exploits	0.98352	0.610389	0.650545	0.864893	0.742513
Fuzzers	0.98352	0.610389	0.746045	0.604409	0.667615
Generic	0.98352	0.610389	0.997497	0.987327	0.992386
Reconnaissance	0.98352	0.610389	0.921843	0.787112	0.849157
Shellcode	0.98352	0.610389	0.878125	0.906155	0.891304
Worms	0.98352	0.610389	0.622240	0.474127	0.536058
Normal	0.98352	0.610389	0.996349	0.998032	0.997190

Por Nó

	Acurácia_Binário	Acurácia_Multiclasse	Acurácia_Balanceada_Binário	Acurácia_Balanceada_Multiclasse
Mean	0.99645	0.983481	0.991431	0.594525

	Precision_Mean	Recall_Mean	F1_Mean
Normal	0.997781	0.998155	0.997968
Attack	0.987262	0.984706	0.985982

Por Nó Pai

	Acurácia_Binário	Acurácia_Multiclasse	Acurácia_Balanceada_Binário	Acurácia_Balanceada_Multiclasse
Mean	0.996447	0.983456	0.991381	0.587681

	Precision_Mean	Recall_Mean	F1_Mean
Normal	0.997765	0.998168	0.997966
Attack	0.987347	0.984595	0.985968

Por Nível

	Acurácia_Binário	Acurácia_Multiclasse	Acurácia_Balanceada_Binário	Acurácia_Balanceada_Multiclasse
Mean	0.996447	0.983427	0.991381	0.58719

	Precision_Mean	Recall_Mean	F1_Mean
Normal	0.997765	0.998168	0.997966
Attack	0.987347	0.984595	0.985968

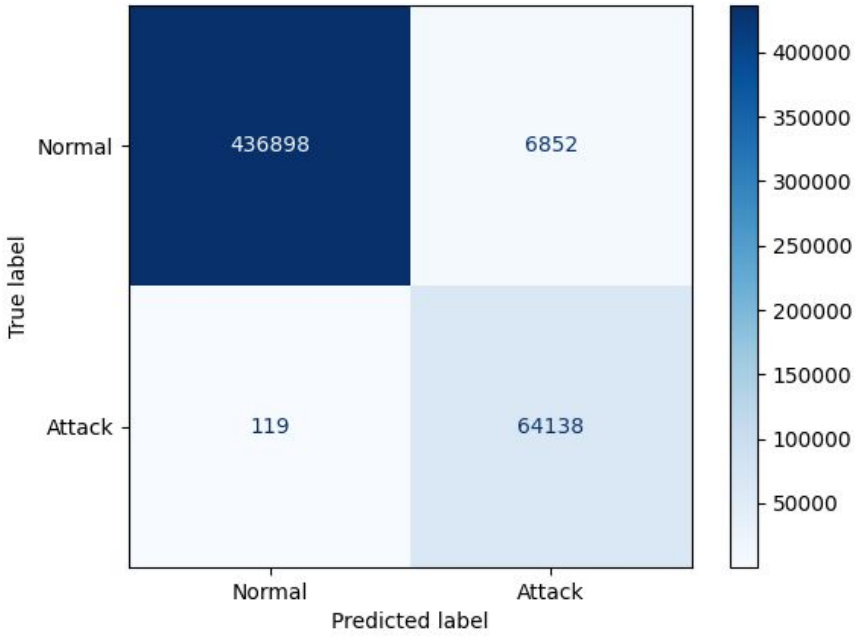
	Precision_Mean	Recall_Mean	F1_Mean
Analysis	0.817582	0.127165	0.219782
Backdoors	0.843748	0.103787	0.184486
DoS	0.312674	0.239655	0.271189
Exploits	0.630403	0.824973	0.714655
Fuzzers	0.757627	0.699358	0.727214
Generic	0.997945	0.987798	0.992845
Reconnaissance	0.928426	0.779149	0.847249
Shellcode	0.868891	0.900768	0.884078
Worms	0.564286	0.284444	0.376370
Normal	0.997781	0.998155	0.997968

	Precision_Mean	Recall_Mean	F1_Mean
Analysis	0.851156	0.126527	0.219837
Backdoors	0.891321	0.098019	0.176129
DoS	0.314806	0.242599	0.273835
Exploits	0.630253	0.823923	0.714164
Fuzzers	0.757163	0.698749	0.726685
Generic	0.997898	0.987429	0.992635
Reconnaissance	0.928417	0.781137	0.848419
Shellcode	0.845588	0.900579	0.871759
Worms	0.603175	0.219683	0.319201
Normal	0.997765	0.998168	0.997966

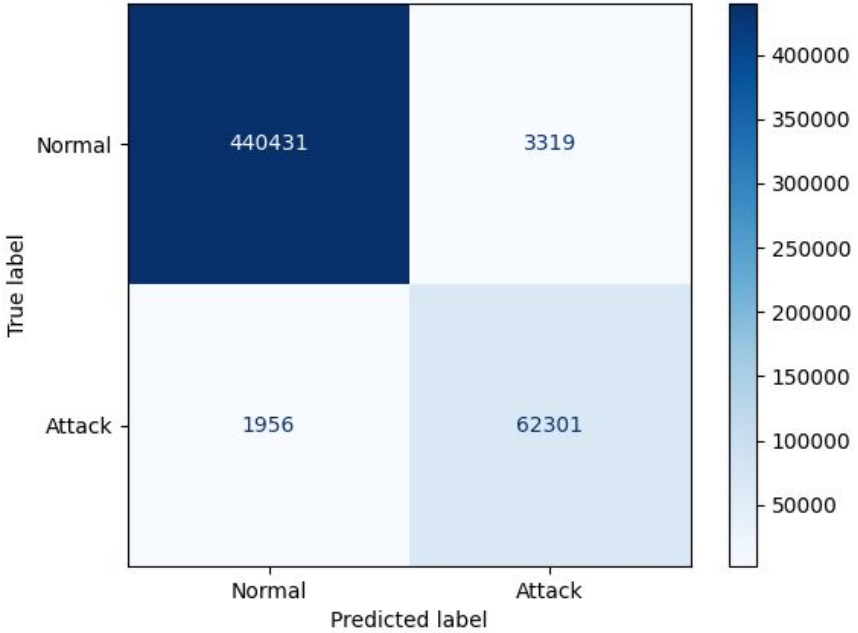
Por Nó, Por Nó Pai e Por Nível

	Precision_Mean	Recall_Mean	F1_Mean
Analysis	0.824467	0.126545	0.218778
Backdoors	0.817051	0.097931	0.174462
DoS	0.309394	0.235067	0.267005
Exploits	0.628746	0.826129	0.714019
Fuzzers	0.761097	0.697845	0.728005
Generic	0.997990	0.987326	0.992630
Reconnaissance	0.925832	0.780283	0.846838
Shellcode	0.841262	0.902922	0.870382
Worms	0.644286	0.219683	0.323480
Normal	0.997765	0.998168	0.997966

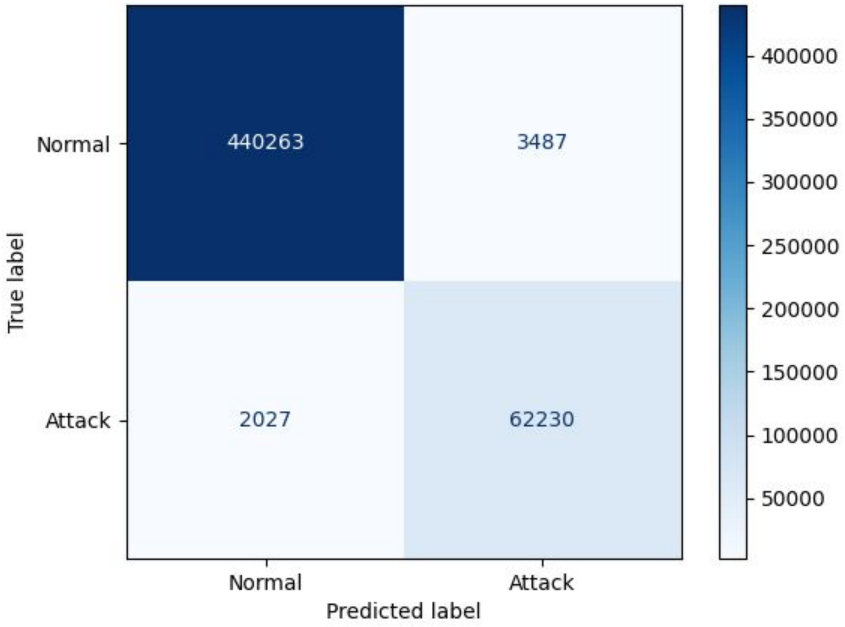
Bagging

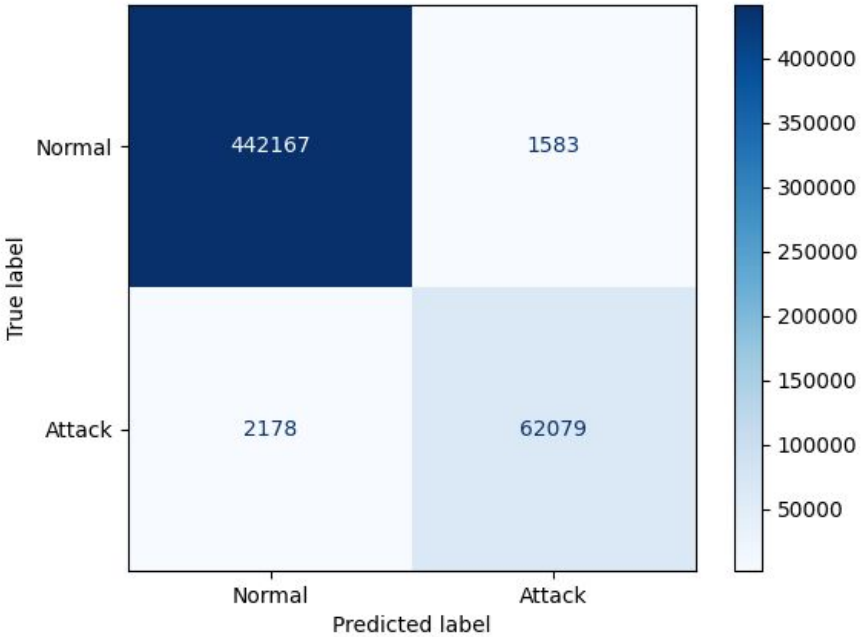


Adaptive Boosting



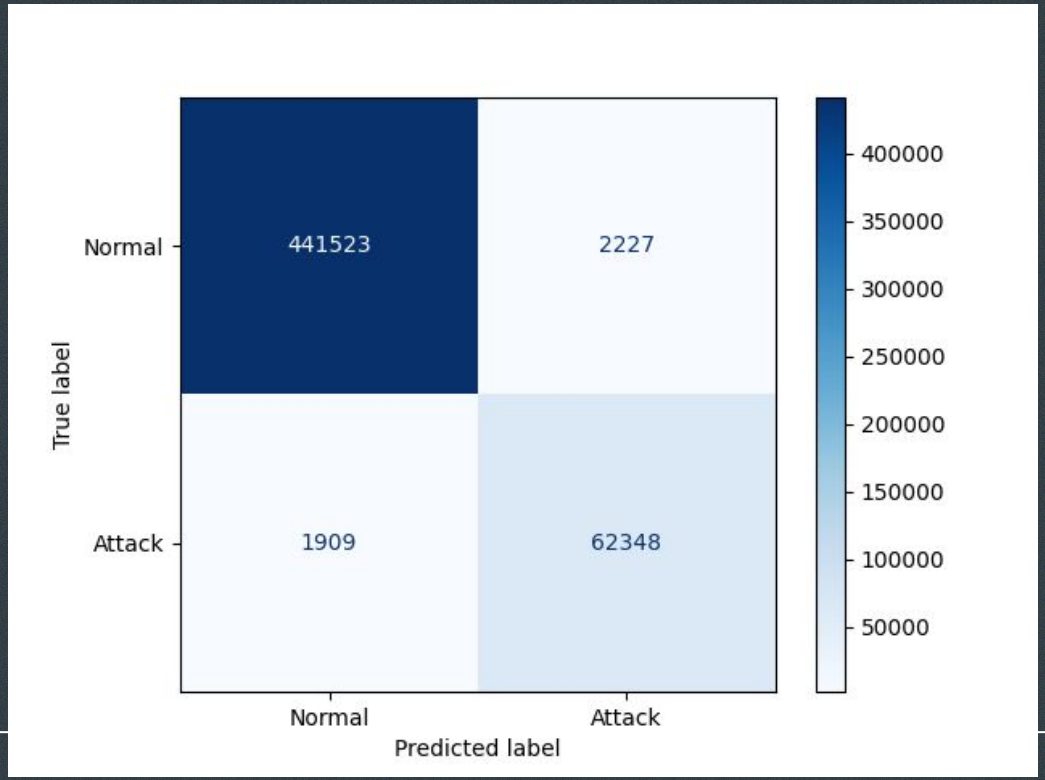
MLP



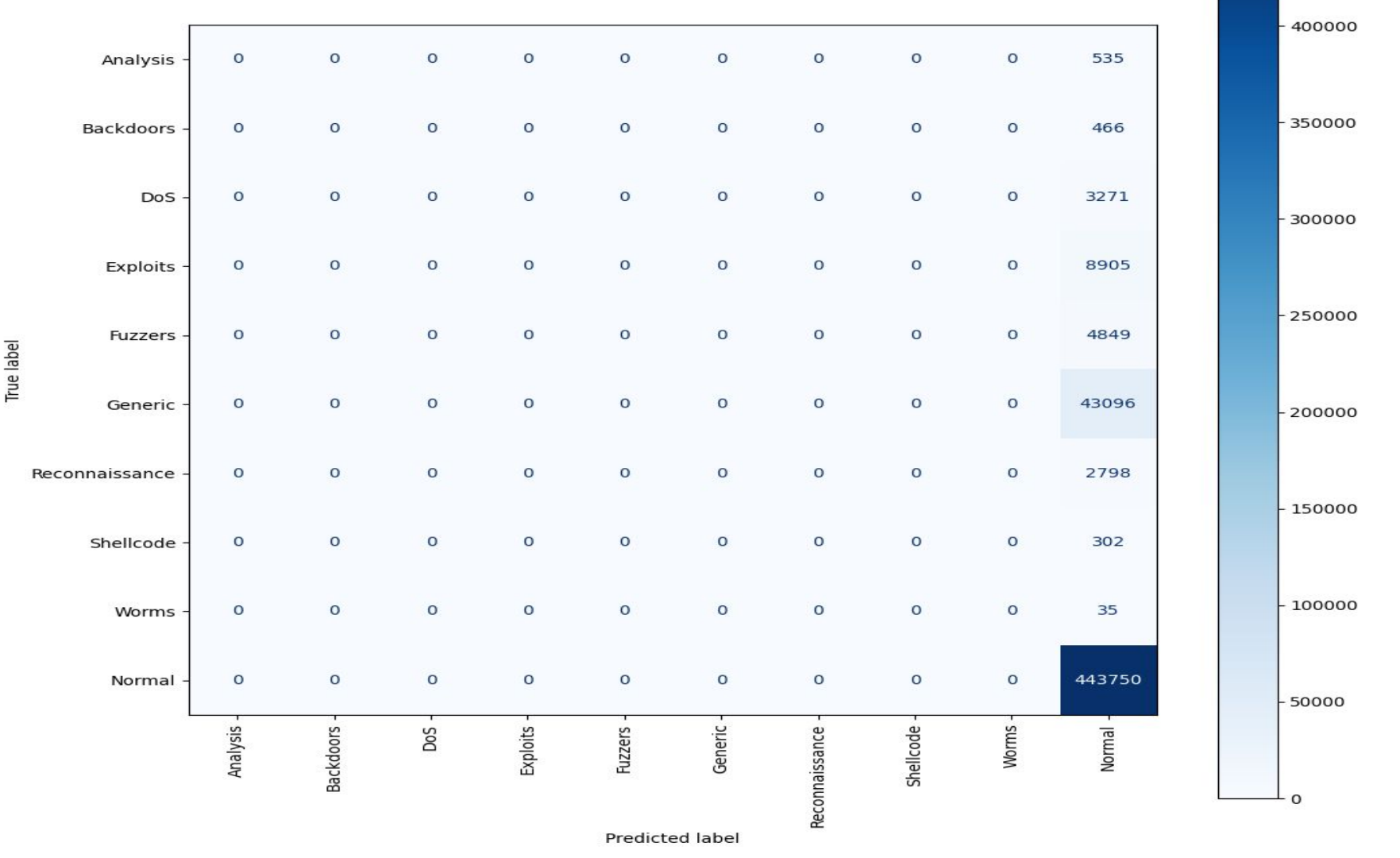


Random Forest

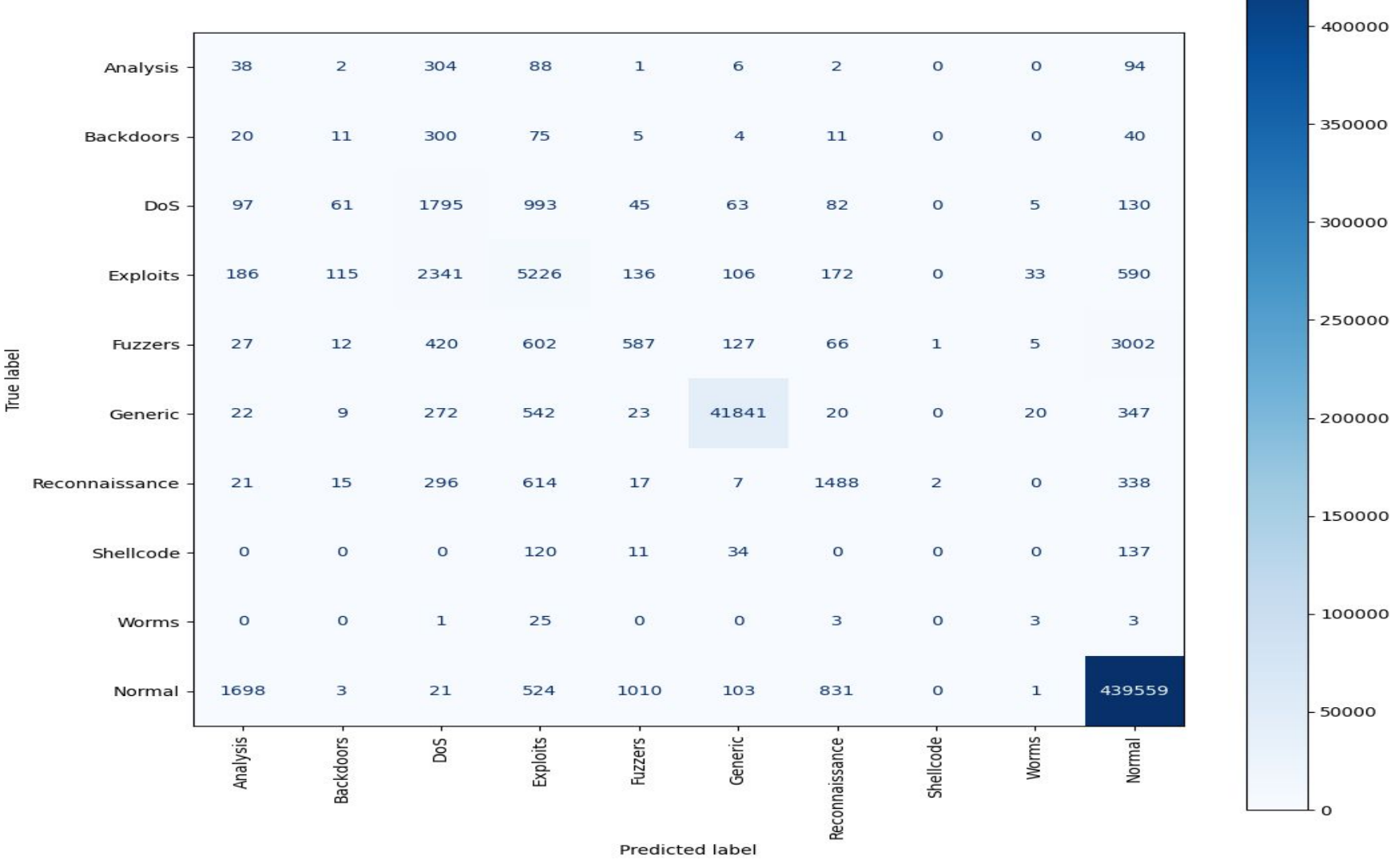
Extreme Gradient Boosting



Bagging


Analysis 0 0 0 0 0 0 0 0 0 535
Backdoors 0 0 0 0 0 0 0 0 0 466
DoS 0 0 0 0 0 0 0 0 0 3271
Exploits 0 0 0 0 0 0 0 0 0 8905
Fuzzers 0 0 0 0 0 0 0 0 0 4849
Generic 0 0 0 0 0 0 0 0 0 43096
Reconnaissance 0 0 0 0 0 0 0 0 0 2798
Shellcode 0 0 0 0 0 0 0 0 0 302
Worms 0 0 0 0 0 0 0 0 0 35
Normal 0 0 0 0 0 0 0 0 0 443750
True label
Analysis
Backdoors
DoS
Exploits
Fuzzers
Generic
Reconnaissance
Shellcode
Worms
Normal
Predicted label
400000
350000
300000
250000
200000
150000
100000
50000
0

Adaptive Boosting



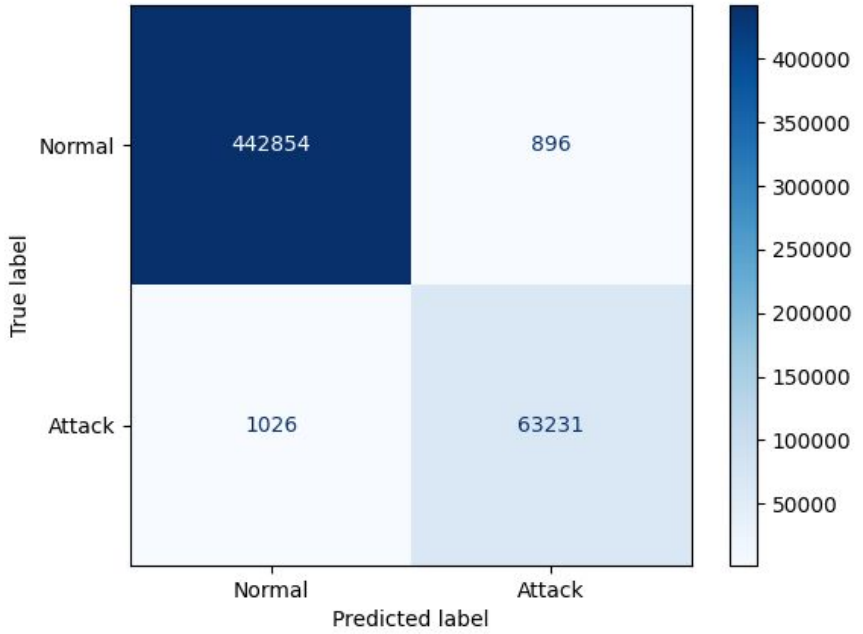
MLP

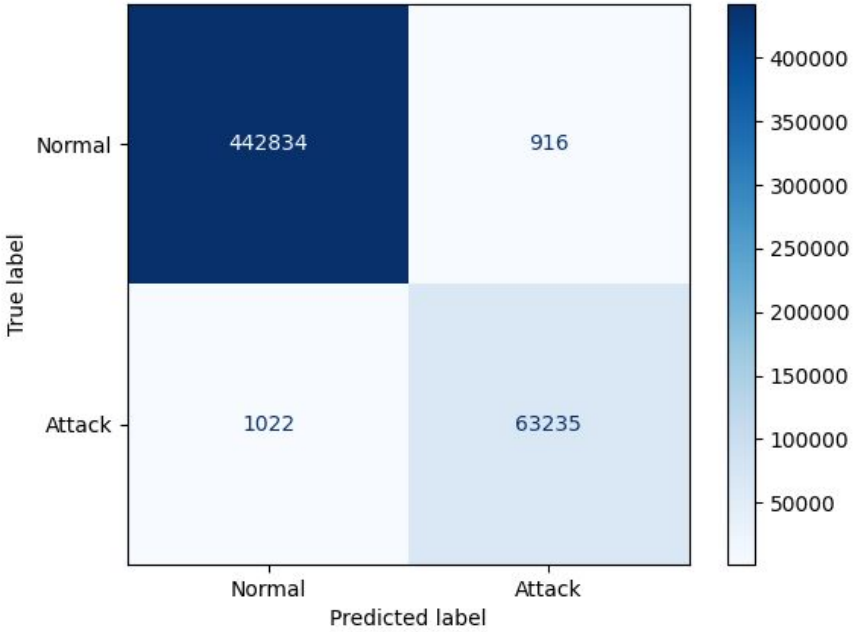

Analysis 21 0 304 20 71 0 0 0 0 119
Backdoors 3 8 300 46 75 6 11 2 0 15
DoS 7 3 2417 578 112 29 46 25 0 54
Exploits 11 1 2924 5231 220 126 124 39 0 229
Fuzzers 10 0 341 100 2327 15 72 22 0 1962
Generic 7 2 360 266 21 42374 18 9 4 35
Reconnaissance 0 0 430 287 25 3 1994 3 0 56
Shellcode 0 0 1 33 10 3 76 154 0 25
Worms 0 0 0 26 1 0 1 2 2 3
Normal 8 0 9 208 830 18 151 20 0 442506
True label
Analysis
Backdoors
DoS
Exploits
Fuzzers
Generic
Reconnaissance
Shellcode
Worms
Normal
Predicted label
400000
350000
300000
250000
200000
150000
100000
50000
0

Random Forest

Extreme Gradient Boosting

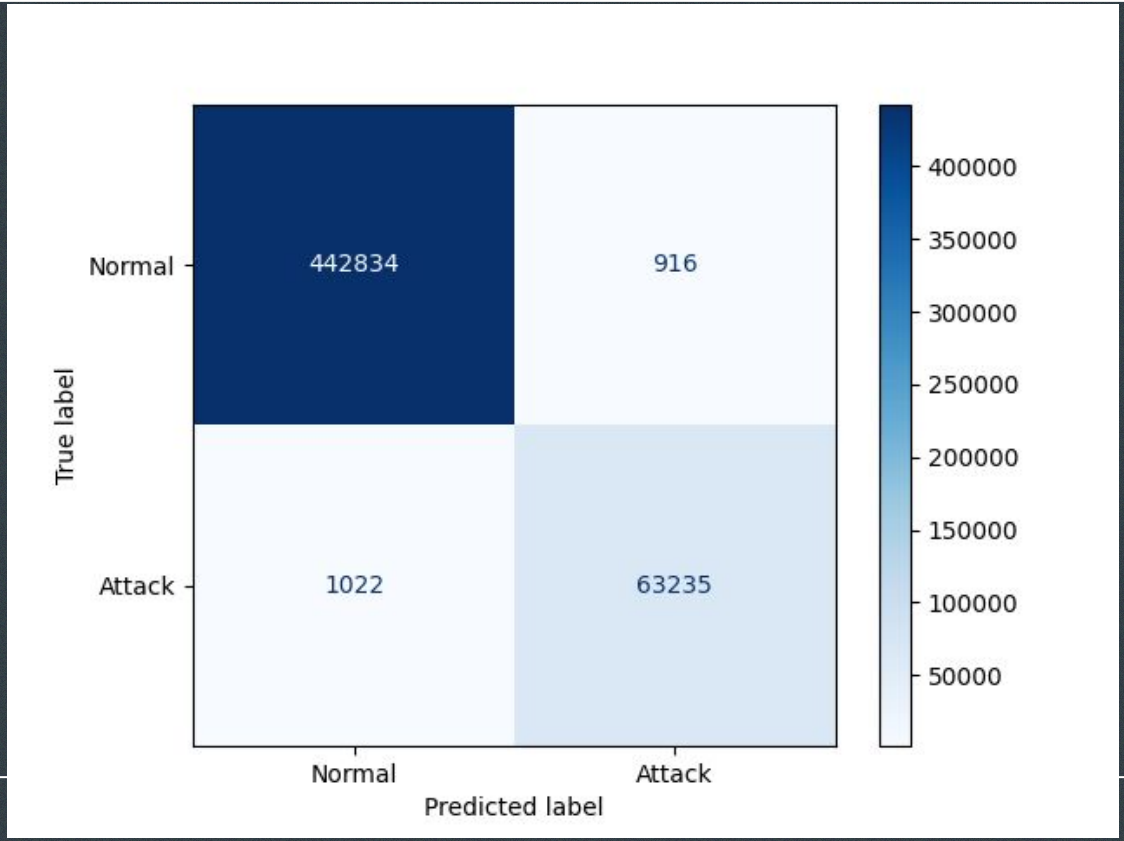
Por Nó



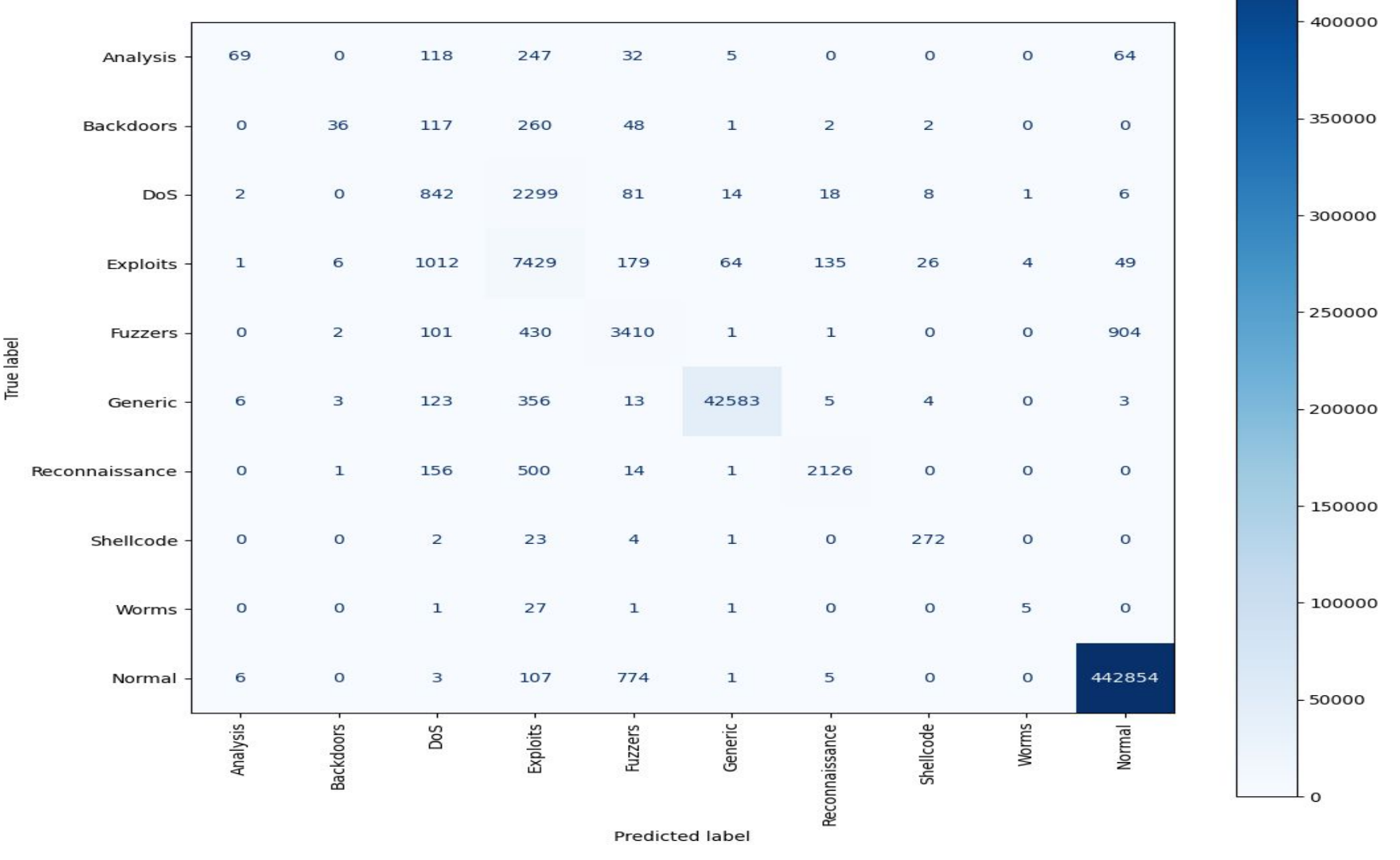


Por Nó Pai

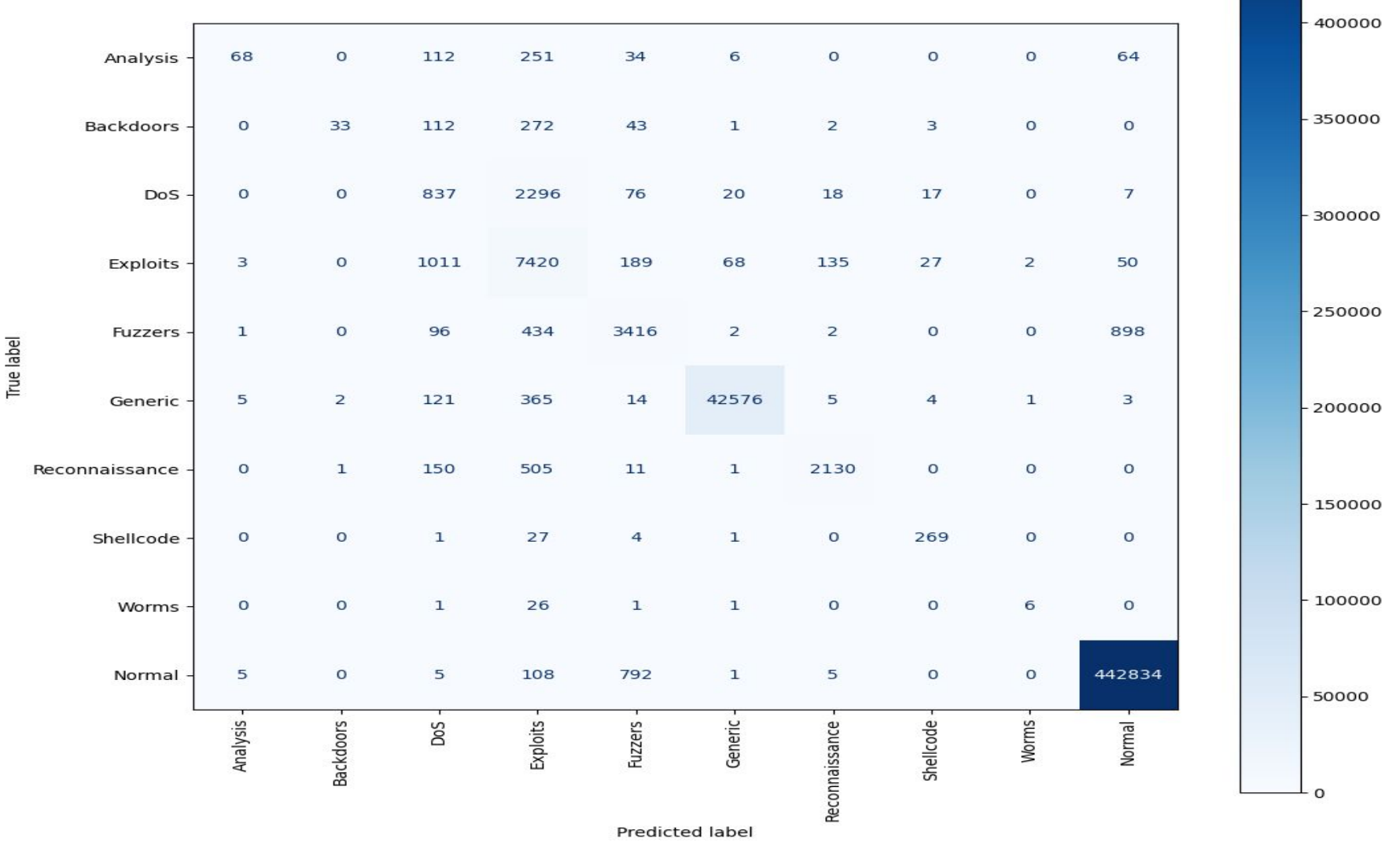
Por Nível


Normal
Attack
True label
442834
916
1022
63235
Normal
Attack
Predicted label
400000
350000
300000
250000
200000
150000
100000
50000

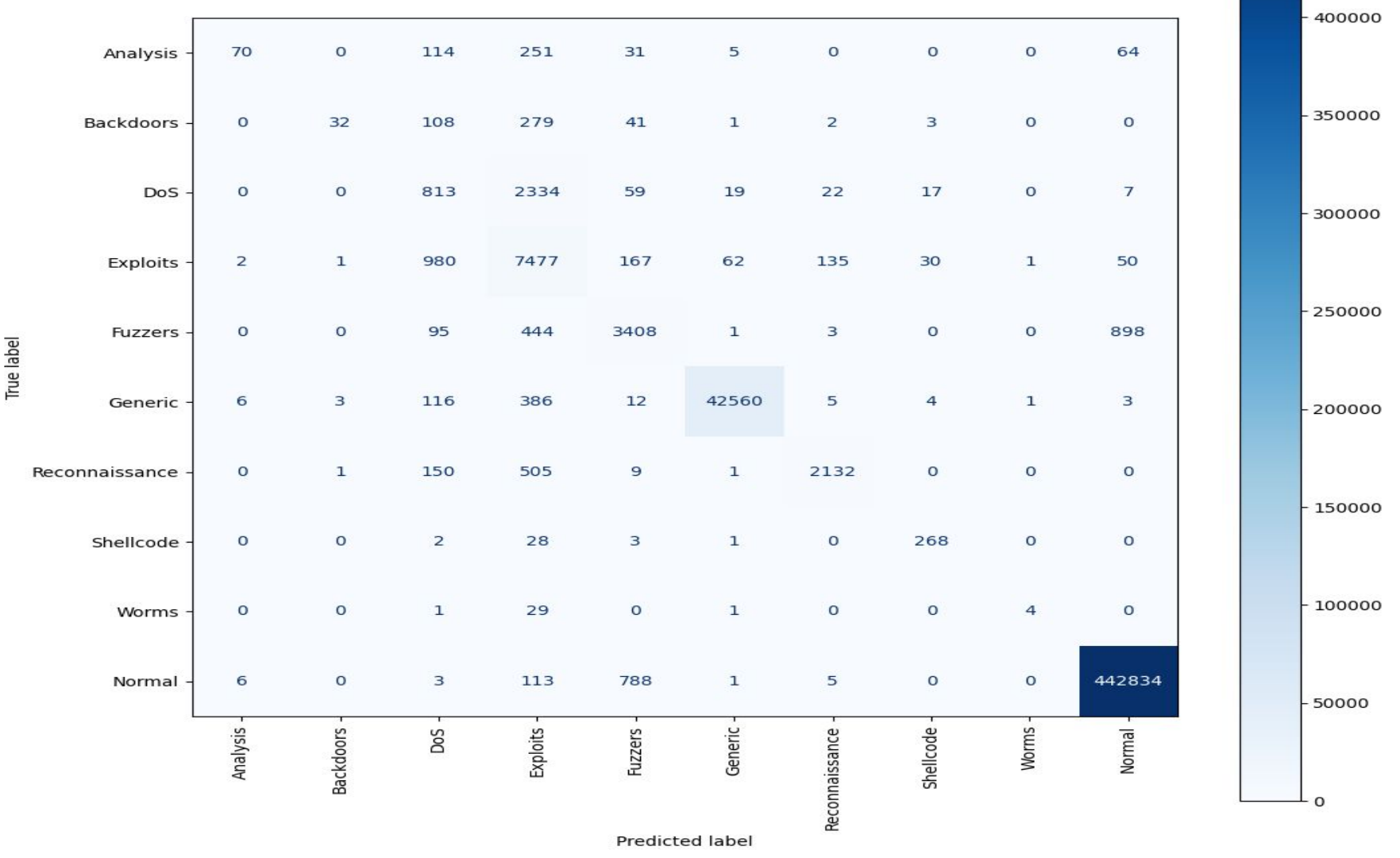
Por Nó


Analysis 69 0 118 247 32 5 0 0 0 64
Backdoors 0 36 117 260 48 1 2 2 0 0
DoS 2 0 842 2299 81 14 18 8 1 6
Exploits 1 6 1012 7429 179 64 135 26 4 49
Fuzzers 0 2 101 430 3410 1 1 0 0 904
Generic 6 3 123 356 13 42583 5 4 0 3
Reconnaissance 0 1 156 500 14 1 2126 0 0 0
Shellcode 0 0 2 23 4 1 0 272 0 0
Worms 0 0 1 27 1 1 0 0 5 0
Normal 6 0 3 107 774 1 5 0 0 442854
True label
Analysis
Backdoors
DoS
Exploits
Fuzzers
Generic
Reconnaissance
Shellcode
Worms
Normal
Predicted label
400000
350000
300000
250000
200000
150000
100000
50000
0

Por Nó Pai


Analysis 68 0 112 251 34 6 0 0 0 64
Backdoors 0 33 112 272 43 1 2 3 0 0
DoS 0 0 837 2296 76 20 18 17 0 7
Exploits 3 0 1011 7420 189 68 135 27 2 50
Fuzzers 1 0 96 434 3416 2 2 0 0 898
Generic 5 2 121 365 14 42576 5 4 1 3
Reconnaissance 0 1 150 505 11 1 2130 0 0 0
Shellcode 0 0 1 27 4 1 0 269 0 0
Worms 0 0 1 26 1 1 0 0 6 0
Normal 5 0 5 108 792 1 5 0 0 442834
True label
Analysis
Backdoors
DoS
Exploits
Fuzzers
Generic
Reconnaissance
Shellcode
Worms
Normal
Predicted label
400000
350000
300000
250000
200000
150000
100000
50000
0

Por Nível


Analysis 70 0 114 251 31 5 0 0 0 64
Backdoors 0 32 108 279 41 1 2 3 0 0
DoS 0 0 813 2334 59 19 22 17 0 7
Exploits 2 1 980 7477 167 62 135 30 1 50
Fuzzers 0 0 95 444 3408 1 3 0 0 898
Generic 6 3 116 386 12 42560 5 4 1 3
Reconnaissance 0 1 150 505 9 1 2132 0 0 0
Shellcode 0 0 2 28 3 1 0 268 0 0
Worms 0 0 1 29 0 1 0 0 4 0
Normal 6 0 3 113 788 1 5 0 0 442834
True label
Analysis
Backdoors
DoS
Exploits
Fuzzers
Generic
Reconnaissance
Shellcode
Worms
Normal
Predicted label
400000
350000
300000
250000
200000
150000
100000
50000
0

Conclusão

- Modelos mais satisfatórios:
  Extreme Gradient Boosting,
  Random Forest e
  Hierárquicos

Obrigado!

Você tem alguma pergunta?

Referências:

- CHANG, Brittany. One of the biggest US insurance companies reportedly paid hackers $40 million ransom after a cyberattack. Business Insider. Dispon´ıvel em: https://www.businessinsider.com/cna-financialhackers-40-million-ransom- cyberattack-2021-5
- AXELSSON, Stefan. Intrusion Detection Systems: A Survey and Taxonomy. CiteSeer. Abril, 2000.
- RAKSHE, Tushar. GONJARI, Vishal. Anomaly based Network Intrusion Detection using Machine Learning Techniques. International Journal of Engineering Research & Technology (IJERT). Maio, 2017.
- AKASHI SATO, Felipe Yuzo. Análise do Desempenho de Algoritmos Classificadores para Detecção de Anomalias em Sistemas de Detecção de Intrusão. Universidade Tecnol´ogica Federal do Paraná (UTFPR). Junho, 2023.
- Moustafa, Nour, and Jill Slay. ”UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set).”Military Communications and Information Systems Conference (MilCIS), 2015. IEEE, 2015.

Referências:

- Moustafa, Nour, and Jill Slay. ”The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set.”Information Security Journal: A Global Perspective (2016): 1-14.
- MIRANDA, Fabio M. KO¨ HNECKE, Niklas. RENARD, Bernhard Y. HiClass: a Python Library for Local Hierarchical Classification Compatible with Scikit-learn. Janeiro, 2023.