Resultados com Fontes Heterogêneas

Lucas Albano

UFMG

Parâmetros dos Experimentos

- Ordinal Patterns
  - dx = 3, taux = 1, janela de 5% do dataset
- Atributos do tráfego de rede, logs do apache e firewall
  - Somente anterior ao ataque
- Seleção de features
  - RFC, LASSO, XGBoost, PCA, ANOVA, Relieff, MultiSURF, Information Gain, Pearson Correlation
  - ⅓ do dataset, shuffle = true
- One Class SVM
  - kernel = poly, nu = 0.05
  - ⅓ do dataset, shuffle = false

Features

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5282 entries, 0 to 5281
Data columns (total 77 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   total.pckts                     5282 non-null   int64
 1   tcp.pct                         5282 non-null   float64
 2   udp.pct                         5282 non-null   float64
 3   http.pct                        5282 non-null   float64
 4   other.pct                       5282 non-null   float64
 5   ip.src.unique                   5282 non-null   int64
 6   ip.dest.unique                  5282 non-null   int64
 7   eth.src.unique                  5282 non-null   int64
 8   eth.dest.unique                 5282 non-null   int64
 9   pkt.length.mean                 5282 non-null   float64
 10  pkt.length.std                  5282 non-null   float64
 11  pkt.length.var                  5282 non-null   float64
 12  pkt.length.min                  5282 non-null   float64
 13  pkt.length.max                  5282 non-null   float64
 14  pkt.length.pquartil             5282 non-null   float64
 15  pkt.length.med                  5282 non-null   float64
 16  pkt.length.tquartil             5282 non-null   float64
 17  pkt.delta.mean                  5282 non-null   float64
 18  pkt.delta.std                   5282 non-null   float64
 19  pkt.delta.var                   5282 non-null   float64
 20  pkt.delta.min                   5282 non-null   float64
 21  pkt.delta.max                   5282 non-null   float64
 22  pkt.delta.pquartil              5282 non-null   float64
 23  pkt.delta.med                   5282 non-null   float64
 24  pkt.delta.tquartil              5282 non-null   float64
 25  ttl.mean                        5282 non-null   float64
 26  ttl.std                         5282 non-null   float64
 27  ttl.var                         5282 non-null   float64
 28  ttl.min                         5282 non-null   float64
 29  ttl.max                         5282 non-null   float64
 30  ttl.pquartil                    5282 non-null   float64
 31  ttl.med                         5282 non-null   float64
 32  ttl.tquartil                    5282 non-null   float64
 33  tcp.flag.urg.count              5282 non-null   int64
 34  tcp.flag.ack.count              5282 non-null   int64
 35  tcp.flag.push.count             5282 non-null   int64
 36  tcp.flag.reset.count            5282 non-null   int64
 37  tcp.flag.syn.count              5282 non-null   int64
 38  tcp.flag.fyn.count              5282 non-null   int64
 39  dest.port.entropy               5282 non-null   float64
 40  src.port.entropy                5282 non-null   float64
 41  tcp.seq.entropy                 5282 non-null   float64
 42  http.statusok                   5282 non-null   float64
 43  http.statusnotok                5282 non-null   float64
 44  http.get.pct                    5282 non-null   float64
 45  http.post.pct                   5282 non-null   float64
 46  http.othermethods.pct           5282 non-null   float64
 47  http.useragent.unique           5282 non-null   int64
 48  http.url.unique                 5282 non-null   int64
 49  http.avg.url.len                5282 non-null   float64
 50  apache.total.access.logs        5282 non-null   float64
 51  apache.unique.ips               5282 non-null   float64
 52  apache.get.count                5282 non-null   float64
 53  apache.nonget.count             5282 non-null   float64
 54  apache.unique.urls              5282 non-null   float64
 55  apache.avg.url.length           5282 non-null   float64
 56  apache.status.200.count         5282 non-null   float64
 57  apache.status.non200.count      5282 non-null   float64
 58  apache.avg.request.size         5282 non-null   float64
 59  apache.unique.user.agents       5282 non-null   float64
 60  firewall.total.logs             5282 non-null   float64
 61  firewall.unique.src.ips         5282 non-null   float64
 62  firewall.unique.dest.ips        5282 non-null   float64
 63  firewall.unique.src.ports       5282 non-null   float64
 64  firewall.unique.dest.ports      5282 non-null   float64
 65  firewall.unique.prot            5282 non-null   float64
 66  firewall.avg.pkt.length         5282 non-null   float64
 67  firewall.avg.ttl                5282 non-null   float64
 68  firewall.avg.win.size           5282 non-null   float64
 69  firewall.count.seq              5282 non-null   float64
 70  firewall.count_ACK              5282 non-null   float64
 71  firewall.unique.marks           5282 non-null   float64
 72  firewall.unique.in.interface    5282 non-null   float64
 73  firewall.unique.mac             5282 non-null   float64
 74  firewall.tos                    5282 non-null   float64
 75  firewall.precedence             5282 non-null   float64
 76  label                           5282 non-null   int64
dtypes: float64(63), int64(14)
memory usage: 3.1 MB
```

Resultados

Algoritmo	Features	Acurácia	Precisão	Recall	F1-Score	Predição
PCA	7	96,24%	97,4%	96,24%	96,59%	21m e 48s
MultiSURF	12	80,88%	95,31%	80,88%	85,66%	21m e 54s
XGBoost	37	87,35%	88,32%	87,35%	87,83%	28m e 1s
Information Gain	1	86,37%	95,45%	86,37%	89,41%	21m e 49s
Relieff	222	94,09%	94,44%	94,09%	91,31%	11m e 3s
RFC	166	85,95%	87,92%	85,95%	86,92%	11m e 4s
ANOVA	20	94%	88,37%	94%	91,1%	N/A
LASSO	9	94%	88,37%	94%	91,1%	N/A
Pearson Correlation	9	94%	88,37%	94%	91,1%	N/A

Algoritmo	Features	Acurácia	Precisão	Recall	F1-Score	Predição
PCA	10	96,18%	97,34%	96,18%	96,54%	21m e 47s
MultiSURF	5	91,52%	89,27%	91,52%	90,33%	21m e 31s
XGBoost	30	80,61%	90,68%	80,61%	84,83%	28m e 1s
Information Gain	1	86,37%	95,45%	86,37%	89,41%	21m e 49s
Relieff	110	93,58%	88,74%	93,58%	90,94%	11m e 1s
RFC	142	90,69%	88,25%	90,69%	89,45%	11m e 4s
ANOVA	91	84,58 %	87,82%	84,58%	86,17%	11m e 1s
LASSO	9	94%	88,37%	94%	91,1%	N/A
Pearson Correlation	9	94%	88,37%	94%	91,1%	N/A

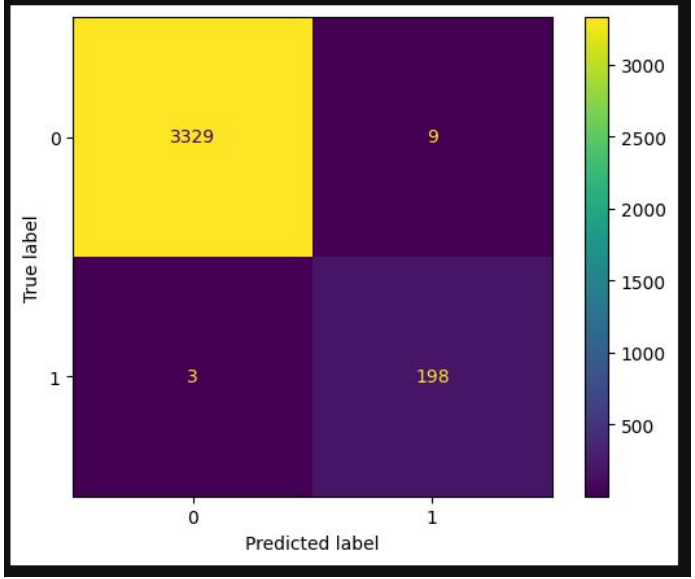
Algoritmo	Features	Acurácia	Precisão	Recall	F1-Score	Predição
PCA	71	89,88%	88,82%	89,88%	89,34%	21m e 42s
MultiSURF	65	90,87%	89,35%	90,87%	90,08%	21m e 55s
XGBoost	24	91,67%	88,94%	91,67%	90,23%	21m e 43s
Information Gain	69	92,12%	89,07%	92,12%	90,48%	21m e 42s
Relieff	57	91,58%	89%	91,58%	90,22%	21m e 42s
RFC	61	91,52%	88,98%	91,52%	90,19%	21m e 42s
ANOVA	70	91,35%	88,94%	91,35%	90,09%	21m e 42s
LASSO	40	90,72%	88,83%	90,72%	89,74%	21m e 42s
Pearson Correlation	71	90,81%	89,45%	90,81%	90,11%	21m e 55s

Algoritmo	Features	Acurácia	Precisão	Recall	F1-Score	Predição
PCA	65	93,85%	89,24%	93,85%	91,08%	21m e 53s
MultiSURF	15	94%	88,37%	94%	91,1%	N/A
XGBoost	40	94,03%	94,39%	94,03%	91,17%	21m e 53s
Information Gain	13	94%	88,37%	94%	91,1%	N/A
Relieff	13	94%	88,37%	94%	91,1%	N/A
RFC	9	94%	88,37%	94%	91,1%	N/A
ANOVA	18	88,51%	95,8%	88,51%	90,91%	21m e 53s
LASSO	36	94,03%	94,39%	94,03%	91,17%	21m e 53s
Pearson Correlation	23	94%	88,37%	94%	91,1%	N/A

MultiSURF - 41 Features

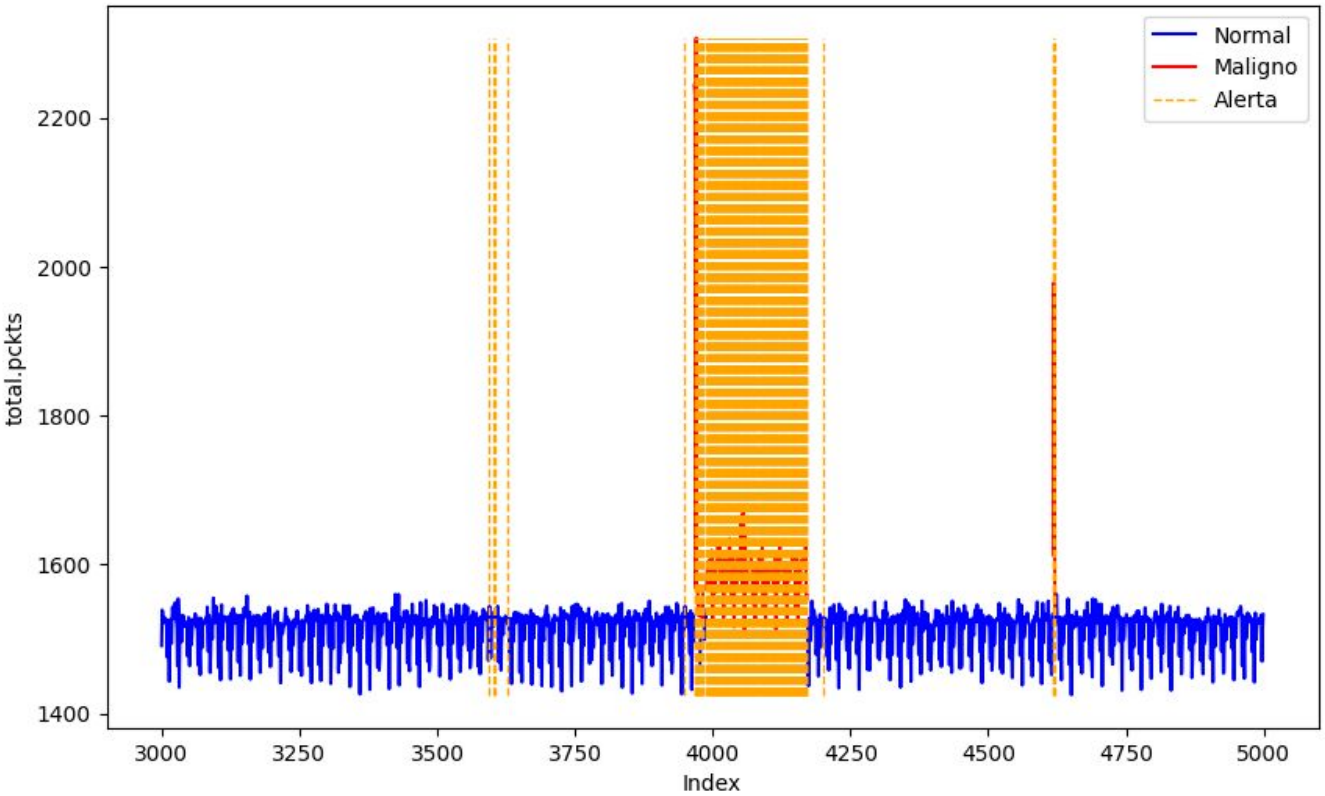
	Feature	Importance
0	apache.unique.ips	0.003251
1	firewall.unique.mac	0.001530
2	firewall.unique.src.ips	0.001530
3	firewall.unique.dest.ips	0.001530
4	firewall.unique.src.ports	0.001530
5	firewall.unique.in.interface	0.001530
6	firewall.unique.prot	0.001530
7	firewall.precedence	0.000574
8	firewall.count_ACK	0.000574
9	firewall.avg.pkt.length	0.000574
10	firewall.tos	0.000574
11	firewall.avg.win.size	0.000574
12	firewall.unique.dest.ports	0.000574
13	firewall.total.logs	0.000574
14	firewall.count.seq	0.000574
15	eth.src.unique	0.000567
16	tcp.flag.urg.count	0.000191
17	ttl.tquartil	0.000000
18	ttl.med	0.000000
19	ttl.pquartil	0.000000
20	firewall.avg.ttl	0.000000
21	ttl.max	0.000000
22	pkt.length.med	0.000000
23	pkt.length.pquartil	0.000000
24	pkt.length.tquartil	0.000000
25	udp.pct	0.000000
26	ttl.min	-0.000515
27	other.pct	-0.000564
28	ttl.mean	-0.000570
29	tcp.flag.reset.count	-0.000574
30	pkt.length.min	-0.000574
31	firewall.unique.marks	-0.000574
32	ttl.var	-0.000574
33	ttl.std	-0.000574
34	ip.src.unique	-0.001094
35	apache.nonget.count	-0.002238
36	http.post.pct	-0.002257
37	eth.dest.unique	-0.003198
38	ip.dest.unique	-0.003198
39	apache.avg.request.size	-0.004323
40	pkt.length.max	-0.007383
41	http.useragent.unique	-0.012588

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3338
1	0.96	0.99	0.97	201
accuracy			1.00	3539
macro avg	0.98	0.99	0.98	3539
weighted avg	1.00	1.00	1.00	3539



28m e 1s

Gráfico


Normal
Maligno
Alerta
total.pckts
2200
2000
1800
1600
1400
3000
3250
3500
3750
4000
4250
4500
4750
5000
Index

Contato:
linkedin.com/in/lucasaoc/
lucasalbano@dcc.ufmg.br

UFMG