

UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Phishing Websites

Daniel Pimenta Furtado



Introdução

Exploração de Dados

Seleção de atributos

Classificação

Visualização dos Resultados - Grupo (Tree)

Conclusão

Introdução

Exploração de Dados

Seleção de atributos

Classificação

Visualização dos Resultados - Grupo (Tree)

Conclusão

Introdução

- ▶ Dataset Phishing Websites Data Set
- ▶ Classificação: Phishing (-1) ou Legítimo (1)
- ▶ 30 atributos (Heurísticas) - Phishing (-1), Suspeito (0) e Legítimo (1)
- ▶ 4 Grupos: Características baseadas na barra de endereço, Características baseadas em anormalidades, Características baseadas em HTML e JavaScript e Características baseadas no domínio

Introdução

Características baseadas na barra de endereço

having_IP_Address: Avalia se um endereço de IP é utilizado como alternativa ao nome de domínio na URL, como “http://125.98.3.123/fake.html”

- ▶ Se a parte de domínio possui endereço IP (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas na barra de endereço

Shortning_Service: Avalia se a URL é proveniente de serviços de encurtamento de URL, essas URL utilizando "HTTP Redirect" para direcionar o usuário para site de destino

- ▶ TinyURL (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas na barra de endereço

having_At_Symbol: Avalia a utilização do símbolo "@", porque quando utilizado o navegador ignora tudo que precede o símbolo na URL

- ▶ Possui o símbolo "@" (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas na barra de endereço

double_slash_redirecting: Avalia a utilização do símbolo "//", porque sua função é de redirecionar o usuário para outro site

- ▶ Posição do último "/" > 7 (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas na barra de endereço

Prefix_Suffix: Avalia a utilização do símbolo "-", esse símbolo é raramente utilizado em sites legítimos e pode ser usado para mascarar a URL, colocando parte de um domínio legítimo. Por exemplo: "http://www.Confirme-paypal.com/"

- ▶ URL possui o símbolo "-" (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas na barra de endereço

having_Sub_Domain: Avalia o número de subdomínios além do country-code top-level domain (ccTLD)

- ▶ Número de pontos na parte do domínio = 1 (Legítimo)
- ▶ Número de pontos na parte do domínio = 2 (Suspeito)
- ▶ Caso contrário (Phishing)

Introdução

Características baseadas na barra de endereço

SSLfinal_State: Avalia a idade do certificado de HTTPS do site e seu emissor

- ▶ Utiliza HTTPS e o emissor é confiável e a idade do certificado ≥ 1 ano (Legítimo)
- ▶ Utiliza HTTPS e o emissor não é confiável (Suspeito)
- ▶ Caso contrário (Phishing)

Introdução

Características baseadas na barra de endereço

Domain_registration_length: Avalia quando o registro do domínio se expira, pois sites maliciosos duram por um curto período de tempo

- ▶ Registro do domínio expira em ≤ 1 ano (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas na barra de endereço

Favicon: Avalia se o site utiliza um favicon que não é carregado a partir do mesmo domínio

- ▶ Favicon carregado a partir de um domínio externo (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas na barra de endereço

port: Avalia se algum serviço possui sua porta fora do valor padrão

- ▶ Porta fora do número padrão (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas na barra de endereço

HTTPS_token: Avalia a presença de um token HTTPS para enganar o usuário. Por exemplo: "http://https-www-paypal-it-webapps-mpp-home.soft-hair.com/"

- ▶ Utiliza token HTTPS (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas em anormalidades

Request_URL: Avalia se os objetos externos contidos no interior do site como imagens, vídeos e sons são carregados partir de domínios externos (Request URL)

- ▶ % de Request URL $< 22\%$ (Legítimo)
- ▶ % de Request URL $\geq 22\%$ e $\leq 61\%$ (Suspeito)
- ▶ Caso contrário (Phishing)

Introdução

Características baseadas em anormalidades

URL_of_Anchor: Avalia se as tags `<a>` no site (Anchor) possuem links para domínios externos ou não possuem link (``)

- ▶ % do URL of Anchor $< 31\%$ (Legítimo)
- ▶ % do URL of Anchor $\geq 31\%$ e $\leq 67\%$ (Suspeito)
- ▶ Caso contrário (Phishing)

Introdução

Características baseadas em anormalidades

Links_in_tags: Avalia a presença de links nas tags `<meta>`, `<script>` e `<link>` (Tags)

- ▶ % de links nas Tags $< 17\%$ (Legítimo)
- ▶ % de links nas Tags $\geq 17\%$ e $\leq 81\%$ (Suspeito)
- ▶ Caso contrário (Phishing)

Introdução

Características baseadas em anormalidades

SFH: Avalia se o Server Form Handler (SFH) possui empty string ou "about:blank" ou é referente a outro domínio da URL original

- ▶ SFH possui empty string ou "about:blank" (Phishing)
- ▶ SFH é referente a outro domínio (Suspeito)
- ▶ Caso Contrário (Legítimo)

Introdução

Características baseadas em anormalidades

Submitting_to_email: Avalia a presença das funções que redireciona informações para algum e-mail (mail(), mailto:)

- ▶ Usando as funções "mail()" ou "mailto:" para submeter informações do usuário (Phishing)
- ▶ Caso Contrário (Legítimo)

Introdução

Características baseadas em anormalidades

Abnormal_URL: Avalia se o host name é parte da URL

- ▶ O host name não é incluído na URL (Phishing)
- ▶ Caso contrário (Legítimo)

Introdução

Características baseadas em HTML e JavaScript

Redirect: Avalia o número de redirecionamento de um site

- ▶ Número de redirecionamento ≤ 1 (Legítimo)
- ▶ Número de redirecionamento ≥ 2 e < 4 (Suspeito)
- ▶ Caso contrário (Phishing)

Introdução

Características baseadas em HTML e JavaScript

on_mouseover: Avalia a utilização do evento "onMouseOver" para esconder uma falsa URL

- ▶ onMouseOver muda a barra de status (Phishing)
- ▶ onMouseOver não muda a barra de status (Legítimo)

Introdução

Características baseadas em HTML e JavaScript

RightClick: Avalia se o site utiliza JavaScript para desabilitar o RightClick

- ▶ RightClick desabilitado (Phishing)
- ▶ Caso Contrário (Legítimo)

Introdução

Características baseadas em HTML e JavaScript

popUpWidnow: Avalia se janelas pop-up possuem campos de textos

- ▶ Janela Pop-up contém campos de texto (Phishing)
- ▶ Caso Contrário (Legítimo)

Introdução

Características baseadas em HTML e JavaScript

Iframe: Avalia se o site utiliza a tag HTML "iframe" para esconder sites adicionais

- ▶ Usando iframe (Phishing)
- ▶ Caso Contrário (Legítimo)

Introdução

Características baseadas no domínio

age_of_domain: Avalia a idade do domínio

- ▶ Idade do domínio ≥ 6 meses (Legítimo)
- ▶ Caso Contrário (Phishing)

Introdução

Características baseadas no domínio

DNSRecord: Avalia se o DNS record é vazio ou não é encontrado

- ▶ Nenhum DNS record para o domínio (Phishing)
- ▶ Caso Contrário (Legítimo)

Introdução

Características baseadas no domínio

web_traffic: Avalia o Website Rank (Alexa database) do site

- ▶ Website Rank < 100000 (Legítimo)
- ▶ Website Rank > 100000 (Suspeito)
- ▶ Não possui Website Rank (Phishing)

Introdução

Características baseadas no domínio

Page_Rank: Avalia o PageRank do site que varia no intervalo de 0 a 1

- ▶ $\text{PageRank} < 0.2$ (Phishing)
- ▶ Caso Contrário (Legítimo)

Introdução

Características baseadas no domínio

Google_Index: Avalia se a página possui Google Index

- ▶ Página possui Google Index (Legítimo)
- ▶ Caso Contrário (Phishing)

Introdução

Características baseadas no domínio

Links_pointing_to_page: Avalia o número de links apontados para o site

- ▶ Link apontando para o site = 0 (Phishing)
- ▶ Link apontando para o site > 0 e ≤ 2 (Suspeito)
- ▶ Caso Contrário (Legítimo)

Introdução

Características baseadas no domínio

Statistical_report: Avalia se o host e o domínio encontram-se entre os sites com piores reputações de acordo com os reports PhishTank e StopBadware

- ▶ Possui reputação ruim no PhishTank ou StopBadware (Phishing)
- ▶ Caso Contrário (Legítimo)

Introdução

Métricas

Acurácia, erro absoluto, precisão (P), recall (R), F1-score (F1) e o coeficiente de matthews (MCC). O F1-Score é definido como:

$$F1 = 2 \frac{PR}{P + R}$$

O coeficiente de matthews (MCC) é definido como:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Onde TP, FP, FN e TN são os *True Positive*, *False Positive*, *False Negative* e *True Negative*

Introdução

Exploração de Dados

Seleção de atributos

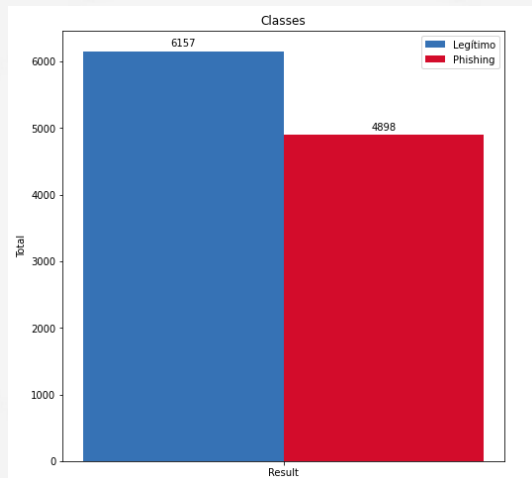
Classificação

Visualização dos Resultados - Grupo (Tree)

Conclusão

Exploração de Dados

O dataset possui 11055 sites



Exploração de Dados

Características baseadas na barra de endereço

Tabela: Distribuição dos dados - Grupo (Características baseadas na barra de endereço)

	-1	0	1
having_IP_Address	3793	0	7262
URL_Length	8960	135	1960
Shortining_Service	1444	0	9611
having_At_Symbol	1655	0	9400
double_slash_redirecting	1429	0	9626
Prefix_Suffix	9590	0	1465
having_Sub_Domain	3363	3622	4070
SSLfinal_State	3557	1167	6331
Domain_registration_length	7398	0	3666
Favicon	2053	0	9002
port	1502	0	9553
HTTPS_token	1796	0	9259

Exploração de Dados

Características baseadas na barra de endereço

Tabela: Distribuição dos dados - Grupo (Características baseadas na barra de endereço)

	-1	0	1
having_IP_Address	3793	0	7262
URL_Length	8960	135	1960
Shortining_Service	1444	0	9611
having_At_Symbol	1655	0	9400
double_slash_redirecting	1429	0	9626
Prefix_Suffix	9590	0	1465
having_Sub_Domain	3363	3622	4070
SSLfinal_State	3557	1167	6331
Domain_registration_length	7398	0	3666
Favicon	2053	0	9002
port	1502	0	9553
HTTPS_token	1796	0	9259

Exploração de Dados

Características baseadas na barra de endereço

Tabela: Classificação - Grupo (Características baseadas na barra de endereço)

	Acurácia	F1	MCC
having_IP_Address	0.5623	0.6394	0.0942
URL_Length	0.4856	0.3035	0.0616
Shortining_Service	0.5193	0.6630	-0.0680
having_At_Symbol	0.5587	0.6864	0.0529
double_slash_redirecting	0.5294	0.6703	-0.0386
Prefix_Suffix	0.5756	0.3844	0.3486
having_Sub_Domain	0.6856	0.7361	0.3606
SSLfinal_State	0.8779	0.9032	0.7388
Domain_registration_length	0.3752	0.2969	-0.2258
Favicon	0.5357	0.6614	-0.0003
port	0.5539	0.6861	0.0364
HTTPS_token	0.5238	0.6585	-0.0399

Exploração de Dados

Características baseadas na barra de endereço

Tabela: Classificação - Grupo (Características baseadas na barra de endereço)

	Acurácia	F1	MCC
having_IP_Address	0.5623	0.6394	0.0942
URL_Length	0.4856	0.3035	0.0616
Shortining_Service	0.5193	0.6630	-0.0680
having_At_Symbol	0.5587	0.6864	0.0529
double_slash_redirecting	0.5294	0.6703	-0.0386
Prefix_Suffix	0.5756	0.3844	0.3486
having_Sub_Domain	0.6856	0.7361	0.3606
SSLfinal_State	0.8779	0.9032	0.7388
Domain_registration_length	0.3752	0.2969	-0.2258
Favicon	0.5357	0.6614	-0.0003
port	0.5539	0.6861	0.0364
HTTPS_token	0.5238	0.6585	-0.0399

Exploração de Dados

Características baseadas na barra de endereço

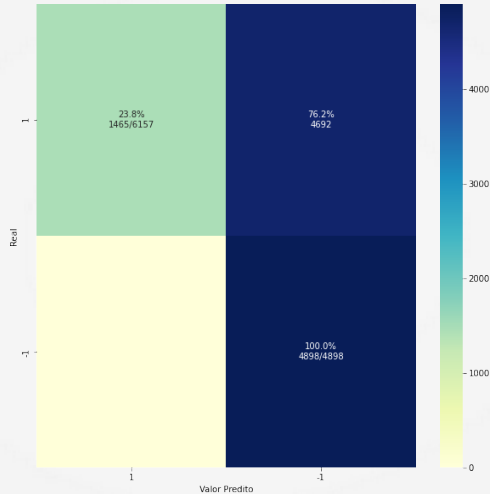


Figura: Matriz de Confusão - Prefix_Suffix

Exploração de Dados

Características baseadas em anormalidades

Tabela: Distribuição dos dados - Grupo (Características baseadas em anormalidade)

	-1	0	1
Request_URL	4495	0	6560
URL_of_Anchor	3282	5337	2436
Links_in_tags	3956	4449	2650
SFH	8440	761	1854
Submitting_to_email	2014	0	9041
Abnormal_URL	1629	0	9426

Exploração de Dados

Características baseadas em anormalidades

Tabela: Classificação - Grupo (Características baseadas em anormalidade)

	Acurácia	F1	MCC
Request_URL	0.6343	0.6821	0.2534
URL_of_Anchor	0.9675	0.9609	0.9339
Links_in_tags	0.6464	0.6172	0.3081
SFH	0.5532	0.3879	0.2224
Submitting_to_email	0.5432	0.6677	0.0182
Abnormal_URL	0.5189	0.6587	-0.0605

Exploração de Dados

Características baseadas em anormalidades

Tabela: Classificação - Grupo (Características baseadas em anormalidade)

	Acurácia	F1	MCC
Request_URL	0.6343	0.6821	0.2534
URL_of_Anchor	0.9675	0.9609	0.9339
Links_in_tags	0.6464	0.6172	0.3081
SFH	0.5532	0.3879	0.2224
Submitting_to_email	0.5432	0.6677	0.0182
Abnormal_URL	0.5189	0.6587	-0.0605

Exploração de Dados

Características baseadas em HTML e JavaScript

Tabela: Distribuição dos dados - Grupo (Características baseadas em HTML e JavaScript)

	-1	0	1
Redirect	0	9776	1279
on_mouseover	1315	0	9740
RightClick	476	0	10579
popUpWidnow	2137	0	8918
Iframe	1012	0	10043

Exploração de Dados

Características baseadas em HTML e JavaScript

Tabela: Classificação - Grupo (Características baseadas em HTML e JavaScript)

	Acurácia	F1	MCC
Redirect	0.5293	0.6922	-
on__mouseover	0.5569	0.6918	0.0418
RightClick	0.5546	0.7058	0.0127
popUpWidnow	0.5350	0.6590	0.0001
Iframe	0.5455	0.6899	-0.0034

Exploração de Dados

Características baseadas no domínio

Tabela: Distribuição dos dados - Grupo (Características baseadas no domínio)

	-1	0	1
age_of_domain	5189	0	5866
DNSRecord	3443	0	7612
web_traffic	2655	2569	5831
Page_Rank	8201	0	2854
Google_Index	1539	0	9516
Links_pointing_to_page	548	6156	4351
Statistical_report	1550	0	9505

Exploração de Dados

Características baseadas no domínio

Tabela: Classificação - Grupo (Características baseadas no domínio)

	Acurácia	F1	MCC
age_of_domain	0.5637	0.5989	0.1215
DNSRecord	0.5563	0.6438	0.0757
web_traffic	0.7067	0.7765	0.3560
Page_Rank	0.5180	0.4086	0.1046
Google_Index	0.5854	0.7076	0.1290
Links_pointing_to_page	0.5650	0.7073	-0.0360
Statistical_report	0.5685	0.6954	0.0799

Exploração de Dados

Características baseadas no domínio

Tabela: Classificação - Grupo (Características baseadas no domínio)

	Acurácia	F1	MCC
age_of_domain	0.5637	0.5989	0.1215
DNSRecord	0.5563	0.6438	0.0757
web_traffic	0.7067	0.7765	0.3560
Page_Rank	0.5180	0.4086	0.1046
Google_Index	0.5854	0.7076	0.1290
Links_pointing_to_page	0.5650	0.7073	-0.0360
Statistical_report	0.5685	0.6954	0.0799

Introdução

Exploração de Dados

Seleção de atributos

Classificação

Visualização dos Resultados - Grupo (Tree)

Conclusão

Seleção de atributos

- ▶ Grupo 1 (Exploração de Dados) = Prefix_Suffix, having_Sub_Domain, SSLfinal_State, Request_URL, URL_of_Anchor, Links_in_tags, SFH, web_traffic, Page_Rank, Google_Index
- ▶ Grupo 2 (Barra de Endereço) = having_IP_Address, URL_Length, Shortining_Service, having_At_Symbol, double_slash_redirecting, Prefix_Suffix, having_Sub_Domain, SSLfinal_State, Domain_registration_length, Favicon, port, HTTPS_token
- ▶ Grupo 3 (Anormalidades) = Request_URL, URL_of_Anchor, Links_in_tags, SFH, Submitting_to_email, Abnormal_URL

Seleção de atributos

- ▶ Grupo 4 (HTML e JavaScript) = Redirect, on__mouseover, RightClick, popUpWidnow, Iframe
- ▶ Grupo 5 (Domínio) = age__of__domain, DNSRecord, web__traffic, Page__Rank, Google__Index, Links__pointing__to__page, Statistical__report

Seleção de atributos

Feature selection SKlearn. L1 - LinearSVC (C=0.001, penalty="L1") e Tree - ExtraTreesClassifier

- ▶ Grupo 6 (L1) = Prefix_Suffix, having_Sub_Domain, SSLfinal_State, Request_URL, URL_of_Anchor, Links_in_tags, SFH, web_traffic, Google_Index
- ▶ Grupo 7 (Tree) = Prefix_Suffix, having_Sub_Domain, SSLfinal_State, URL_of_Anchor, Links_in_tags, web_traffic

Seleção de atributos

- ▶ Nenhuma das 3 seleções (1,6 e 7) selecionaram atributos do Grupo 4 (HTML e JavaScript)
- ▶ A seleção L1 e da exploração de dados se diferenciam somente no atributo Page_Rank

Seleção de atributos

- ▶ Grupo 8 (Soma): Soma os resultados dos atributos de cada grupo apresentado pelos autores e depois realiza o rescale min-max das quatro somas resultantes
- ▶ Grupo 9: Todos os atributos

Introdução

Exploração de Dados

Seleção de atributos

Classificação

Visualização dos Resultados - Grupo (Tree)

Conclusão

Classificação

Banco de treinamento e teste

- ▶ 80% para treinamento e 20% para teste
- ▶ KFold Cross Validation (`StratifiedKFold(n_splits=5)`)
- ▶ LogisticRegression (Regressão Logística), LinearSVC (SVM Linear) e Random Forest
- ▶ GridSearchCV
- ▶ Fase de teste

Classificação

Banco de treinamento e teste

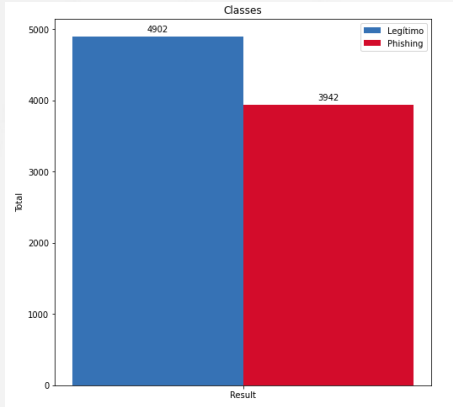


Figura: Base de Treino

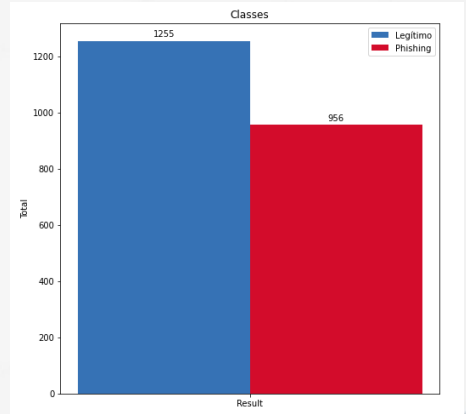


Figura: Base de Teste

Classificação

Treinamento

Tabela: Distribuição por características

	Random Forest	LogisticRegression	LinearSVC
Grupo 1	0.952671	0.930683	0.931840
Grupo 2	0.917702	0.904317	0.905580
Grupo 3	0.890296	0.878645	0.881101
Grupo 4	0.714688	0.703560	0.703560
Grupo 5	0.773944	0.715063	0.715734
Grupo 6	0.948334	0.930729	0.932076
Grupo 7	0.942996	0.923717	0.925873
Grupo 8	0.864812	0.773261	0.773064
Grupo 9	0.972709	0.935974	0.936362

Classificação

Treinamento

Tabela: Resultados Treinamento - F1-score

	Random Forest	LogisticRegression	LinearSVC
Grupo 1	0.952671	0.930683	0.931840
Grupo 2	0.917702	0.904317	0.905580
Grupo 3	0.890296	0.878645	0.881101
Grupo 4	0.714688	0.703560	0.703560
Grupo 5	0.773944	0.715063	0.715734
Grupo 6	0.948334	0.930729	0.932076
Grupo 7	0.942996	0.923717	0.925873
Grupo 8	0.864812	0.773261	0.773064
Grupo 9	0.972709	0.935974	0.936362

Classificação

GridSearchCV

- ▶ `n_estimators = [50,100,200,500]`
- ▶ `max_features = ['auto', 'sqrt']`
- ▶ `max_depth = [None, 10,25,50]`
- ▶ `min_samples_split = [2, 16, 64]`
- ▶ `min_samples_leaf = [1, 16, 64]`
- ▶ `bootstrap = [True, False]`
- ▶ `criterion = ['gini', 'entropy']`

1152 Possibilidades

Classificação

GridSearchCV

- ▶ Grupo 1 = 'bootstrap': True, 'criterion': 'gini', 'max_depth': None, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200
- ▶ Grupo 7 = 'bootstrap': True, 'criterion': 'gini', 'max_depth': 50, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 16, 'n_estimators': 50
- ▶ Grupo 9 = 'bootstrap': False, 'criterion': 'entropy', 'max_depth': None, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100

Classificação

GridSearchCV

Tabela: GridSearchCV

	F1
Grupo 1	0.95346
Grupo 7	0.944187
Grupo 9	0.974146

Classificação

Teste

Tabela: Teste

	F1	Acurácia	MCC
Grupo 1	0.9505	0.9435	0.8847
Grupo 7	0.9491	0.9412	0.8803
Grupo 9	0.9719	0.9679	0.9346

Introdução

Exploração de Dados

Seleção de atributos

Classificação

Visualização dos Resultados - Grupo (Tree)

Conclusão

Visualização dos Resultados - Grupo (Tree)

Treinamento - Percentage Split

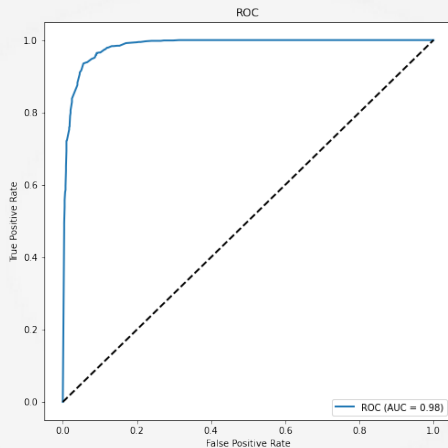


Figura: Curva Roc - Treino (Percentage Split)

Visualização dos Resultados - Grupo (Tree)

Treinamento - Percentage Split

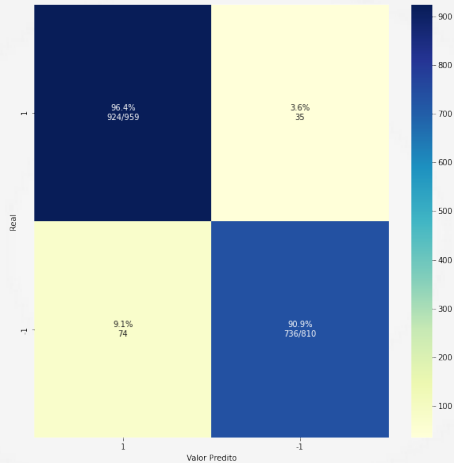


Figura: Matriz de Confusão - Treino (Percentage Split)

Visualização dos Resultados - Grupo (Tree)

Treinamento - Percentage Split

Tabela: Resultados - Treino (Percentage Split)

	Erro Absoluto	Precisão	Recall	F1	Acurácia	MCC
Resultados	0.1232	0.9259	0.9635	0.9443	0.9384	0.8763

Visualização dos Resultados - Grupo (Tree)

Treinamento - KFold Cross Validation

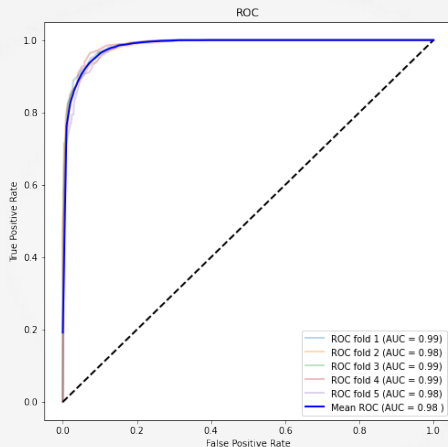


Figura: Curva Roc - Treino (CV)

Visualização dos Resultados - Grupo (Tree)

Treinamento - KFold Cross Validation

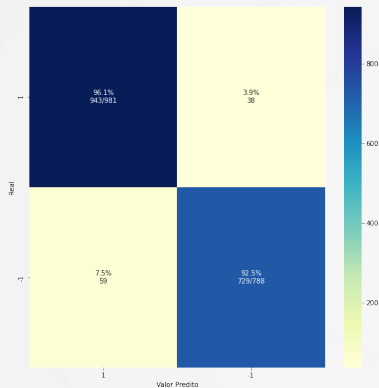


Figura: Matriz de Confusão - Treino (CV1)

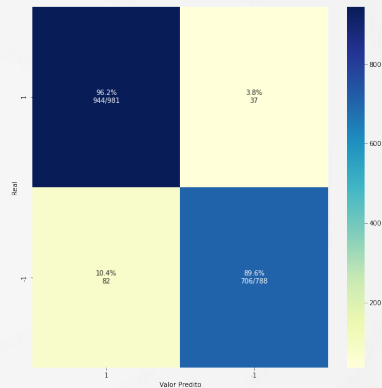


Figura: Matriz de Confusão - Treino (CV2)

Visualização dos Resultados - Grupo (Tree)

Treinamento - KFold Cross Validation

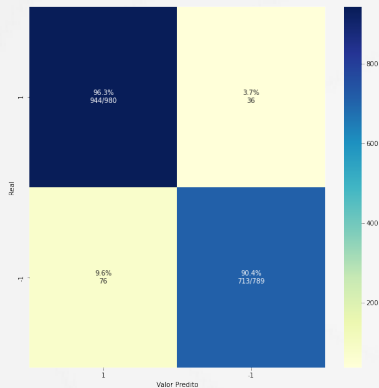


Figura: Matriz de Confusão - Treino (CV3)

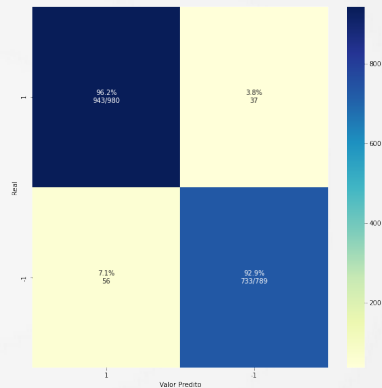


Figura: Matriz de Confusão - Treino (CV4)

Visualização dos Resultados - Grupo (Tree)

Treinamento - KFold Cross Validation

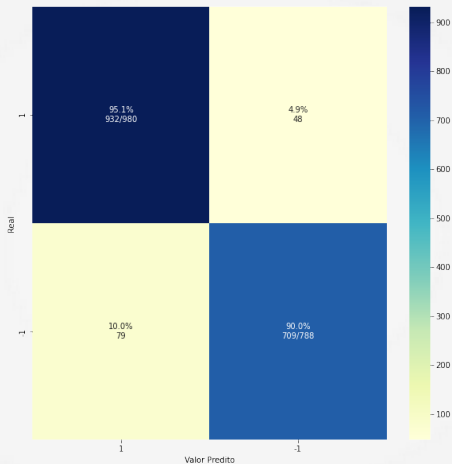


Figura: Matriz de Confusão - Treino (CV5)

Visualização dos Resultados - Grupo (Tree)

Teste

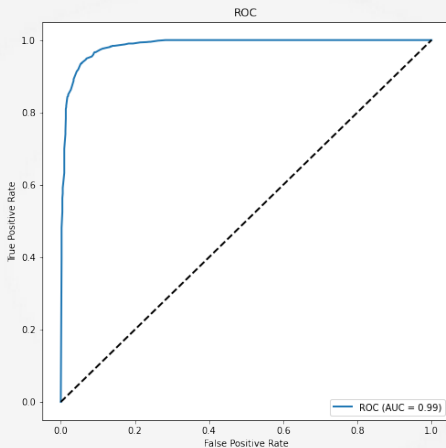


Figura: Curva Roc - Teste

Visualização dos Resultados - Grupo (Tree)

Teste

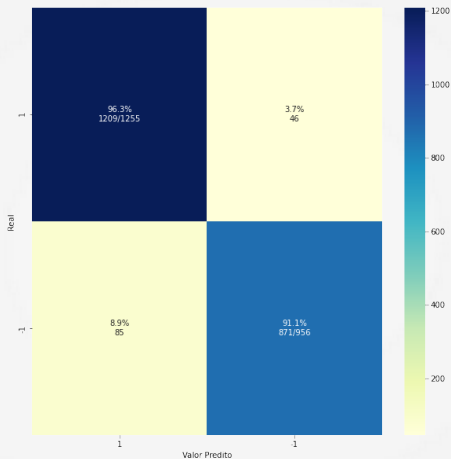


Figura: Matriz de Confusão - Teste

Visualização dos Resultados - Grupo (Tree)

Teste

Tabela: Resultados - Teste

	Erro Absoluto	Precisão	Recall	F1	Acurácia	MCC
Resultados	0.1185	0.9343	0.9633	0.9486	0.9408	0.8793

Introdução

Exploração de Dados

Seleção de atributos

Classificação

Visualização dos Resultados - Grupo (Tree)

Conclusão

Conclusão

- ▶ O Grupo (Tree): Prefix_Suffix, having_Sub_Domain, SSLfinal_State, URL_of_Anchor, Links_in_tags, web_traffic, apresentou resultados satisfatórios
- ▶ O Grupo (HTML e JavaScript) não obteve bons resultados
- ▶ Dataset previamente tratado e anonimizado
- ▶ RandomForest apresentou os melhores resultados vs (LinearSVC/RegressionLogistic)

MUITO OBRIGADO