

```

1 D:\DataScience_and_MachineLearning\Projects\
   Entri_Final_Project\Using_Pycharm\.venv\Scripts\
   python.exe D:\DataScience_and_MachineLearning\
   Projects\Entri_Final_Project\Using_Pycharm\main.py
2
3 Running data_preprocessing.py...
4             FILENAME ... label
5 0          521848.txt ... 1
6 1          31372.txt ... 1
7 2          597387.txt ... 1
8 3          554095.txt ... 1
9 4          151578.txt ... 1
10 ...       ... ...
11 235790    660997.txt ... 1
12 235791    77185.txt ... 1
13 235792    622132.txt ... 1
14 235793    7503962.txt ... 0
15 235794    384822.txt ... 1
16
17 [235795 rows x 56 columns]
18             URLLength DomainLength ...
19 count      235795.000000 235795.000000 ... 235795.
   000000 235795.000000
20 mean       34.573095     21.470396 ... 49.
   262516 0.571895
21 std        41.314153     9.150793 ... 161.
   027430 0.494805
22 min        13.000000     4.000000 ... 0.
   000000 0.000000
23 25%       23.000000     16.000000 ... 1.
   000000 0.000000
24 50%       27.000000     20.000000 ... 10.
   000000 1.000000
25 75%       34.000000     24.000000 ... 57.
   000000 1.000000
26 max        6097.000000    110.000000 ... 27516.
   000000 1.000000
27
28 [8 rows x 51 columns]
29             count           mean

```

29	...	75%	max	
30	URLLength		235795.0	34.573095
	...	34.000000	6.097000e+03	
31	DomainLength		235795.0	21.470396
	...	24.000000	1.100000e+02	
32	IsDomainIP		235795.0	0.002706
	...	0.000000	1.000000e+00	
33	URLSimilarityIndex		235795.0	78.430778
	...	100.000000	1.000000e+02	
34	CharContinuationRate		235795.0	0.845508
	...	1.000000	1.000000e+00	
35	TLDLegitimateProb		235795.0	0.260423
	...	0.522907	5.229071e-01	
36	URLCharProb		235795.0	0.055747
	...	0.062875	9.082366e-02	
37	TLDLength		235795.0	2.764456
	...	3.000000	1.300000e+01	
38	NoOfSubDomain		235795.0	1.164758
	...	1.000000	1.000000e+01	
39	Has0bfuscation		235795.0	0.002057
	...	0.000000	1.000000e+00	
40	NoOf0bfuscatedChar		235795.0	0.024861
	...	0.000000	4.470000e+02	
41	ObfuscationRatio		235795.0	0.000138
	...	0.000000	3.480000e-01	
42	NoOfLettersInURL		235795.0	19.428919
	...	20.000000	5.191000e+03	
43	LetterRatioInURL		235795.0	0.515946
	...	0.594000	9.260000e-01	
44	NoOfDigitsInURL		235795.0	1.881011
	...	0.000000	2.011000e+03	
45	DigitRatioInURL		235795.0	0.028616
	...	0.000000	6.840000e-01	
46	NoOfEqualsInURL		235795.0	0.062241
	...	0.000000	1.760000e+02	
47	NoOfQMarkInURL		235795.0	0.029403
	...	0.000000	4.000000e+00	
48	NoOfAmpersandInURL		235795.0	0.025056
	...	0.000000	1.490000e+02	
49	NoOfOtherSpecialCharsInURL	235795.0		2.340198
	...	3.000000	4.990000e+02	

50	SpacialCharRatioInURL	235795.0	0.063309
	...	0.083000	3.970000e-01
51	IsHTTPS	235795.0	0.782625
	...	1.000000	1.000000e+00
52	LineOfCode	235795.0	1141.900443
	...	1277.000000	4.426660e+05
53	LargestLineLength	235795.0	12789.532624
	...	8047.000000	1.397573e+07
54	HasTitle	235795.0	0.861261
	...	1.000000	1.000000e+00
55	DomainTitleMatchScore	235795.0	50.131427
	...	100.000000	1.000000e+02
56	URLTitleMatchScore	235795.0	52.122098
	...	100.000000	1.000000e+02
57	HasFavicon	235795.0	0.361768
	...	1.000000	1.000000e+00
58	Robots	235795.0	0.266541
	...	1.000000	1.000000e+00
59	IsResponsive	235795.0	0.624513
	...	1.000000	1.000000e+00
60	NoOfURLRedirect	235795.0	0.133438
	...	0.000000	1.000000e+00
61	NoOfSelfRedirect	235795.0	0.040107
	...	0.000000	1.000000e+00
62	HasDescription	235795.0	0.440183
	...	1.000000	1.000000e+00
63	NoOfPopup	235795.0	0.221765
	...	0.000000	6.020000e+02
64	NoOfFrame	235795.0	1.588638
	...	1.000000	1.602000e+03
65	HasExternalFormSubmit	235795.0	0.043987
	...	0.000000	1.000000e+00
66	HasSocialNet	235795.0	0.456570
	...	1.000000	1.000000e+00
67	HasSubmitButton	235795.0	0.414301
	...	1.000000	1.000000e+00
68	HasHiddenFields	235795.0	0.377799
	...	1.000000	1.000000e+00
69	HasPasswordField	235795.0	0.102263
	...	0.000000	1.000000e+00
70	Bank	235795.0	0.127089

70	...	0.000000	1.000000e+00	
71	Pay		235795.0	0.237007
	...	0.000000	1.000000e+00	
72	Crypto		235795.0	0.023474
	...	0.000000	1.000000e+00	
73	HasCopyrightInfo		235795.0	0.486775
	...	1.000000	1.000000e+00	
74	NoOfImage		235795.0	26.075689
	...	29.000000	8.956000e+03	
75	NoOfCSS		235795.0	6.333111
	...	8.000000	3.582000e+04	
76	NoOfJS		235795.0	10.522305
	...	15.000000	6.957000e+03	
77	NoOfSelfRef		235795.0	65.071113
	...	88.000000	2.739700e+04	
78	NoOfEmptyRef		235795.0	2.377629
	...	1.000000	4.887000e+03	
79	NoOfExternalRef		235795.0	49.262516
	...	57.000000	2.751600e+04	
80	label		235795.0	0.571895
	...	1.000000	1.000000e+00	
81				
82	[51 rows x 8 columns]			
83	<class 'pandas.core.frame.DataFrame'>			
84	RangeIndex: 235795 entries, 0 to 235794			
85	Data columns (total 56 columns):			
86	# Column		Non-Null	Count
	Dtype			
87	---	-----	-----	-----

88	0 FILENAME		235795	non-null
	object			
89	1 URL		235795	non-null
	object			
90	2 URLLength		235795	non-null
	int64			
91	3 Domain		235795	non-null
	object			
92	4 DomainLength		235795	non-null
	int64			
93	5 IsDomainIP		235795	non-null

93	int64	
94	6 TLD object	235795 non-null
95	7 URLSimilarityIndex float64	235795 non-null
96	8 CharContinuationRate float64	235795 non-null
97	9 TLDLegitimateProb float64	235795 non-null
98	10 URLCharProb float64	235795 non-null
99	11 TLDLength int64	235795 non-null
100	12 NoOfSubDomain int64	235795 non-null
101	13 Has0bfuscation int64	235795 non-null
102	14 NoOf0bfuscatedChar int64	235795 non-null
103	15 ObfuscationRatio float64	235795 non-null
104	16 NoOfLettersInURL int64	235795 non-null
105	17 LetterRatioInURL float64	235795 non-null
106	18 NoOfDigitsInURL int64	235795 non-null
107	19 DigitRatioInURL float64	235795 non-null
108	20 NoOfEqualsInURL int64	235795 non-null
109	21 NoOfQMarkInURL int64	235795 non-null
110	22 NoOfAmpersandInURL int64	235795 non-null
111	23 NoOfOtherSpecialCharsInURL int64	235795 non-null
112	24 SpacialCharRatioInURL float64	235795 non-null
113	25 IsHTTPS int64	235795 non-null

114	26	LineOfCode	235795	non-null
		int64		
115	27	LargestLineLength	235795	non-null
		int64		
116	28	HasTitle	235795	non-null
		int64		
117	29	Title	235795	non-null
		object		
118	30	DomainTitleMatchScore	235795	non-null
		float64		
119	31	URLTitleMatchScore	235795	non-null
		float64		
120	32	HasFavicon	235795	non-null
		int64		
121	33	Robots	235795	non-null
		int64		
122	34	IsResponsive	235795	non-null
		int64		
123	35	NoOfURLRedirect	235795	non-null
		int64		
124	36	NoOfSelfRedirect	235795	non-null
		int64		
125	37	HasDescription	235795	non-null
		int64		
126	38	NoOfPopup	235795	non-null
		int64		
127	39	NoOfFrame	235795	non-null
		int64		
128	40	HasExternalFormSubmit	235795	non-null
		int64		
129	41	HasSocialNet	235795	non-null
		int64		
130	42	HasSubmitButton	235795	non-null
		int64		
131	43	HasHiddenFields	235795	non-null
		int64		
132	44	HasPasswordField	235795	non-null
		int64		
133	45	Bank	235795	non-null
		int64		
134	46	Pay	235795	non-null

```

134 int64
135 47 Crypto           235795 non-null
    int64
136 48 HasCopyrightInfo 235795 non-null
    int64
137 49 NoOfImage        235795 non-null
    int64
138 50 NoOfCSS          235795 non-null
    int64
139 51 NoOfJS           235795 non-null
    int64
140 52 NoOfSelfRef      235795 non-null
    int64
141 53 NoOfEmptyRef     235795 non-null
    int64
142 54 NoOfExternalRef   235795 non-null
    int64
143 55 label            235795 non-null
    int64
144 dtypes: float64(10), int64(41), object(5)
145 memory usage: 100.7+ MB
146 None
147 FILENAME            0
148 URL                 0
149 URLLength           0
150 Domain              0
151 DomainLength         0
152 IsDomainIP          0
153 TLD                 0
154 URLSimilarityIndex  0
155 CharContinuationRate 0
156 TLDLegitimateProb   0
157 URLCharProb         0
158 TLDLength           0
159 NoOfSubDomain        0
160 Has0bfuscation       0
161 NoOf0bfuscatedChar   0
162 ObfuscationRatio     0
163 NoOfLettersInURL    0
164 LetterRatioInURL     0
165 NoOfDigitsInURL      0

```

```
166 DigitRatioInURL          0
167 NoOfEqualsInURL          0
168 NoOfQMarkInURL          0
169 NoOfAmpersandInURL      0
170 NoOfOtherSpecialCharsInURL 0
171 SpacialCharRatioInURL    0
172 IsHTTPS                  0
173 LineOfCode                0
174 LargestLineLength         0
175 HasTitle                  0
176 Title                     0
177 DomainTitleMatchScore     0
178 URLTitleMatchScore        0
179 HasFavicon                0
180 Robots                    0
181 IsResponsive              0
182 NoOfURLRedirect           0
183 NoOfSelfRedirect           0
184 HasDescription             0
185 NoOfPopup                 0
186 NoOfiFrame                 0
187 HasExternalFormSubmit       0
188 HasSocialNet               0
189 HasSubmitButton             0
190 HasHiddenFields            0
191 HasPasswordField           0
192 Bank                      0
193 Pay                        0
194 Crypto                     0
195 HasCopyrightInfo           0
196 NoOfImage                  0
197 NoOfCSS                    0
198 NoOfJS                     0
199 NoOfSelfRef                 0
200 NoOfEmptyRef                0
201 NoOfExternalRef             0
202 label                      0
203 dtype: int64
204 <class 'pandas.core.frame.DataFrame'>
205 RangeIndex: 235795 entries, 0 to 235794
206 Data columns (total 53 columns):
```

	# Column	Non-Null Count
	Dtype	
207	0 URLLength	235795 non-null
	int64	
208	1 DomainLength	235795 non-null
	int64	
209	2 IsDomainIP	235795 non-null
	int64	
210	3 TLD	235795 non-null
	object	
211	4 URLSimilarityIndex	235795 non-null
	float64	
212	5 CharContinuationRate	235795 non-null
	float64	
213	6 TLDLegitimateProb	235795 non-null
	float64	
214	7 URLCharProb	235795 non-null
	float64	
215	8 TLDLength	235795 non-null
	int64	
216	9 NoOfSubDomain	235795 non-null
	int64	
217	10 Has0bfuscation	235795 non-null
	int64	
218	11 NoOf0bfuscatedChar	235795 non-null
	int64	
219	12 ObfuscationRatio	235795 non-null
	float64	
220	13 NoOfLettersInURL	235795 non-null
	int64	
221	14 LetterRatioInURL	235795 non-null
	float64	
222	15 NoOfDigitsInURL	235795 non-null
	int64	
223	16 DigitRatioInURL	235795 non-null
	float64	
224	17 NoOfEqualsInURL	235795 non-null
	int64	
225	18 NoOfQMarkInURL	235795 non-null

227	int64	
228	19 NoOfAmpersandInURL int64	235795 non-null
229	20 NoOfOtherSpecialCharsInURL int64	235795 non-null
230	21 SpacialCharRatioInURL float64	235795 non-null
231	22 IsHTTPS int64	235795 non-null
232	23 LineOfCode int64	235795 non-null
233	24 LargestLineLength int64	235795 non-null
234	25 HasTitle int64	235795 non-null
235	26 Title object	235795 non-null
236	27 DomainTitleMatchScore float64	235795 non-null
237	28 URLTitleMatchScore float64	235795 non-null
238	29 HasFavicon int64	235795 non-null
239	30 Robots int64	235795 non-null
240	31 IsResponsive int64	235795 non-null
241	32 NoOfURLRedirect int64	235795 non-null
242	33 NoOfSelfRedirect int64	235795 non-null
243	34 HasDescription int64	235795 non-null
244	35 NoOfPopup int64	235795 non-null
245	36 NoOfiFrame int64	235795 non-null
246	37 HasExternalFormSubmit int64	235795 non-null
247	38 HasSocialNet int64	235795 non-null

248	39	HasSubmitButton int64	235795	non-null
249	40	HasHiddenFields int64	235795	non-null
250	41	HasPasswordField int64	235795	non-null
251	42	Bank int64	235795	non-null
252	43	Pay int64	235795	non-null
253	44	Crypto int64	235795	non-null
254	45	HasCopyrightInfo int64	235795	non-null
255	46	NoOfImage int64	235795	non-null
256	47	NoOfCSS int64	235795	non-null
257	48	NoOfJS int64	235795	non-null
258	49	NoOfSelfRef int64	235795	non-null
259	50	NoOfEmptyRef int64	235795	non-null
260	51	NoOfExternalRef int64	235795	non-null
261	52	label int64	235795	non-null
262	dtypes: float64(10), int64(41), object(2)			
263	memory usage: 95.3+ MB			
264	None			
265	0	231		
266	1	254		
267	2	647		
268	3	231		
269	4	503		
270		...		
271	235790	231		
272	235791	647		
273	235792	157		
274	235793	258		

```
275 235794    231
276 Name: TLD, Length: 235795, dtype: int64
277 0          197871
278 1          93363
279 2          187338
280 3          81289
281 4          69782
282      ...
283 235790    160270
284 235791    191644
285 235792    126238
286 235793    134912
287 235794    9826
288 Name: Title, Length: 235795, dtype: int64
289 <class 'pandas.core.frame.DataFrame'>
290 RangeIndex: 235795 entries, 0 to 235794
291 Data columns (total 53 columns):
292 #   Column           Non-Null Count
293 Dtype
294 ---  -----
295 
294 0   URLLength        235795 non-null
294 1   int64
295 1   DomainLength     235795 non-null
295 1   int64
296 2   IsDomainIP       235795 non-null
296 2   int64
297 3   TLD               235795 non-null
297 3   int64
298 4   URLSimilarityIndex 235795 non-null
298 4   float64
299 5   CharContinuationRate 235795 non-null
299 5   float64
300 6   TLDLegitimateProb 235795 non-null
300 6   float64
301 7   URLCharProb       235795 non-null
301 7   float64
302 8   TLDLength         235795 non-null
302 8   int64
303 9   NoOfSubDomain     235795 non-null
303 9   int64
```

304	10	Has0bfuscation	235795	non-null
		int64		
305	11	No0bfuscatedChar	235795	non-null
		int64		
306	12	ObfuscationRatio	235795	non-null
		float64		
307	13	NoOfLettersInURL	235795	non-null
		int64		
308	14	LetterRatioInURL	235795	non-null
		float64		
309	15	NoOfDigitsInURL	235795	non-null
		int64		
310	16	DigitRatioInURL	235795	non-null
		float64		
311	17	NoOfEqualsInURL	235795	non-null
		int64		
312	18	NoOfQMarkInURL	235795	non-null
		int64		
313	19	NoOfAmpersandInURL	235795	non-null
		int64		
314	20	NoOfOtherSpecialCharsInURL	235795	non-null
		int64		
315	21	SpacialCharRatioInURL	235795	non-null
		float64		
316	22	IsHTTPS	235795	non-null
		int64		
317	23	Line0fCode	235795	non-null
		int64		
318	24	LargestLineLength	235795	non-null
		int64		
319	25	HasTitle	235795	non-null
		int64		
320	26	Title	235795	non-null
		int64		
321	27	DomainTitleMatchScore	235795	non-null
		float64		
322	28	URLTitleMatchScore	235795	non-null
		float64		
323	29	HasFavicon	235795	non-null
		int64		
324	30	Robots	235795	non-null

324	int64	
325	31 IsResponsive	235795 non-null
	int64	
326	32 NoOfURLRedirect	235795 non-null
	int64	
327	33 NoOfSelfRedirect	235795 non-null
	int64	
328	34 HasDescription	235795 non-null
	int64	
329	35 NoOfPopup	235795 non-null
	int64	
330	36 NoOfFrame	235795 non-null
	int64	
331	37 HasExternalFormSubmit	235795 non-null
	int64	
332	38 HasSocialNet	235795 non-null
	int64	
333	39 HasSubmitButton	235795 non-null
	int64	
334	40 HasHiddenFields	235795 non-null
	int64	
335	41 HasPasswordField	235795 non-null
	int64	
336	42 Bank	235795 non-null
	int64	
337	43 Pay	235795 non-null
	int64	
338	44 Crypto	235795 non-null
	int64	
339	45 HasCopyrightInfo	235795 non-null
	int64	
340	46 NoOfImage	235795 non-null
	int64	
341	47 NoOfCSS	235795 non-null
	int64	
342	48 NoOfJS	235795 non-null
	int64	
343	49 NoOfSelfRef	235795 non-null
	int64	
344	50 NoOfEmptyRef	235795 non-null
	int64	

```
345 51 NoOfExternalRef          235795 non-null
      int64
346 52 label                  235795 non-null
      int64
347 dtypes: float64(10), int64(43)
348 memory usage: 95.3 MB
349 None
350          URLLength DomainLength ...
      NoOfExternalRef  label
351 0      -0.086486   0.276436 ...
      124      1
352 1      -0.280125   -0.597807 ...
      217      1
353 2      -0.134896   0.057875 ...
      5       1
354 3      -0.207510   -0.269966 ...
      31      1
355 4      -0.038077   0.494997 ...
      85      1
356 ...     ...
      ...
357 235790 -0.134896   0.057875 ...
      191      1
358 235791 -0.159101   -0.051405 ...
      31      1
359 235792 -0.110691   0.167156 ...
      67      1
360 235793  0.494430   2.789885 ...
      0       0
361 235794 -0.038077   0.494997 ...
      261      1
362
363 [235795 rows x 53 columns]
364 data_preprocessing.py executed successfully.
365
366 Running data_visualization.py...
367
368
369
370
371
```

```
372
373
374
375
376
377 data_visualization.py executed successfully.
378
379 Running handle_outliers.py...
380 handle_outliers.py executed successfully.
381
382 Running train_logistic_regression.py...
383          URLLength  DomainLength  ...  NoOfEmptyRef
384 104874   -0.255920      -0.488527  ...      1.0
385 176003   -0.304330      -0.707087  ...      4.0
386 137445   -0.207510      -0.269966  ...      0.0
387 38100    0.107152       1.150679  ...      5.0
388 65088    -0.110691      0.167156  ...      1.0
389 ...
390 ...
391 ...
392 ...
393 ...
394 ...
395
396 [188636 rows x 52 columns]
397          URLLength  DomainLength  ...  NoOfEmptyRef
398 173310   -0.231715      -0.379246  ...      0.0
           NoOfExternalRef
           11.0
```

399	62486	-0.231715 127.0	-0.379246	...	0.0
400	14858	-0.255920 6.0	-0.488527	...	0.0
401	82943	-0.062281 208.0	0.385716	...	2.0
402	72539	-0.280125 9.0	-0.597807	...	0.0
403
		...			
404	109651	-0.207510 0.0	-0.379246	...	0.0
405	133536	-0.352739 45.0	-0.925648	...	1.0
406	192832	-0.086486 19.0	0.276436	...	3.0
407	92195	-0.134896 17.0	0.057875	...	0.0
408	153247	-0.255920 31.0	-0.488527	...	0.0
409					
410		[47159 rows x 52 columns]			
411	104874	1.0			
412	176003	1.0			
413	137445	0.0			
414	38100	1.0			
415	65088	1.0			
416		...			
417	160997	1.0			
418	155379	0.0			
419	174379	0.0			
420	920	1.0			
421	172904	0.0			
422		Name: label, Length: 188636, dtype: float64			
423	173310	1.0			
424	62486	1.0			
425	14858	1.0			
426	82943	1.0			
427	72539	1.0			
428		...			
429	109651	0.0			

```

430 133536    1.0
431 192832    1.0
432 92195     1.0
433 153247    1.0
434 Name: label, Length: 47159, dtype: float64
435 Train set size: (188636, 52)
436 Test set size: (47159, 52)
437 Accuracy: 0.9928115524078118
438 Precision: 0.9957556126438065
439 Recall: 0.9916573971078977
440 F1-Score: 0.9937022794404503
441
442 Confusion Matrix:
443 [[20075  114]
444 [ 225 26745]]
445
446 Classification Report:
447              precision    recall   f1-score
448           support
449          0.0      0.99      0.99      0.99
450          20189
451          26970
452          accuracy      0.99
453          47159
454          macro avg      0.99      0.99      0.99
455          47159
456          weighted avg      0.99      0.99      0.99
457          47159
458
459          Feature  Importance
460          8        TLDLength    1.981998
461          4        URLSimilarityIndex  1.921730
462          27       DomainTitleMatchScore  0.925024
463          14       LetterRatioInURL    0.747684
464          9        NoOfSubDomain    0.653336
465          20       NoOfOtherSpecialCharsInURL  0.645640
466          47       NoOfCSS        0.634204
467          50       NoOfEmptyRef    0.426643

```

465	34	HasDescription	0.418321		
466	5	CharContinuationRate	0.409219		
467	25	HasTitle	0.408198		
468	45	HasCopyrightInfo	0.397569		
469	6	TLDLegitimateProb	0.362407		
470	13	NoOfLettersInURL	0.341651		
471	1	DomainLength	0.339345		
472	38	HasSocialNet	0.325988		
473	0	URLLength	0.295873		
474	31	IsResponsive	0.283261		
475	23	LineOfCode	0.281390		
476	15	NoOfDigitsInURL	0.277271		
477		train_logistic_regression.py	executed successfully.		
478					
479		Running train_Ridge_Lasso.py...			
480		URLLength DomainLength ... NoOfEmptyRef			
		NoOfExternalRef			
481	104874	-0.255920	-0.488527	...	1.0
		28.0			
482	176003	-0.304330	-0.707087	...	4.0
		22.0			
483	137445	-0.207510	-0.269966	...	0.0
		0.0			
484	38100	0.107152	1.150679	...	5.0
		11.0			
485	65088	-0.110691	0.167156	...	1.0
		135.0			
486
		...			
487	160997	-0.062281	0.385716	...	3.0
		8.0			
488	155379	-0.255920	-0.379246	...	0.0
		0.0			
489	174379	0.470225	1.806361	...	0.0
		2.0			
490	920	-0.183305	-0.160685	...	2.0
		63.0			
491	172904	-0.304330	-0.597807	...	0.0
		0.0			
492					
493		[188636 rows x 52 columns]			

		URLLength	DomainLength	...	NoOfEmptyRef
494		NoOfExternalRef			
495	173310	-0.231715 11.0	-0.379246	...	0.0
496	62486	-0.231715 127.0	-0.379246	...	0.0
497	14858	-0.255920 6.0	-0.488527	...	0.0
498	82943	-0.062281 208.0	0.385716	...	2.0
499	72539	-0.280125 9.0	-0.597807	...	0.0
500
		...			
501	109651	-0.207510 0.0	-0.379246	...	0.0
502	133536	-0.352739 45.0	-0.925648	...	1.0
503	192832	-0.086486 19.0	0.276436	...	3.0
504	92195	-0.134896 17.0	0.057875	...	0.0
505	153247	-0.255920 31.0	-0.488527	...	0.0
506					
507		[47159 rows x 52 columns]			
508	104874	1.0			
509	176003	1.0			
510	137445	0.0			
511	38100	1.0			
512	65088	1.0			
513		...			
514	160997	1.0			
515	155379	0.0			
516	174379	0.0			
517	920	1.0			
518	172904	0.0			
519		Name: label, Length: 188636, dtype: float64			
520	173310	1.0			
521	62486	1.0			
522	14858	1.0			

```
523 82943    1.0
524 72539    1.0
525      ...
526 109651   0.0
527 133536   1.0
528 192832   1.0
529 92195    1.0
530 153247   1.0
531 Name: label, Length: 47159, dtype: float64
532 Train set size: (188636, 52)
533 Test set size: (47159, 52)
534 Ridge Regression Performance:
535 MSE: 0.015421835218770706,
536 RMSE: 0.12418468190067045,
537 R2 Score: 0.937010308890109
538
539 Lasso Regression Performance:
540 MSE: 0.021343181102498807,
541 RMSE: 0.146093056311718,
542 R2 Score: 0.9128248768141081
543 Ridge Regression - Classification Metrics:
544 Accuracy: 0.9994,
545 Precision: 0.9990,
546 Recall: 1.0000,
547 F1-score: 0.9995
548
549 Confusion Matrix:
550 [[20161    28]
551 [     0 26970]]
552
553 Lasso Regression - Classification Metrics:
554 Accuracy: 0.9991,
555 Precision: 0.9984,
556 Recall: 1.0000,
557 F1-score: 0.9992
558
559 Confusion Matrix:
560 [[20147    42]
561 [     0 26970]]
562
563 Classification Report:Ridge
```

		precision	recall	f1-score
564	support			
565				
566	0.0	1.00	1.00	1.00
567	20189	1.0	1.00	1.00
568	26970			
569	accuracy			1.00
570	47159	1.00	1.00	1.00
571	macro avg	1.00	1.00	1.00
572	47159			
573	weighted avg	1.00	1.00	1.00
574	47159			
575	Classification Report:Lasso	precision	recall	f1-score
576	support			
577	0.0	1.00	1.00	1.00
578	20189	1.0	1.00	1.00
579	26970			
580	accuracy			1.00
581	47159	1.00	1.00	1.00
582	macro avg	1.00	1.00	1.00
583	47159			
584	Top 20 Features (Ridge):			
585		Feature	Importance	
586	21	SpacialCharRatioInURL	0.788162	
587	0	URLLength	0.674654	
588	16	DigitRatioInURL	0.534370	
589	24	LargestLineLength	0.393614	
590	4	URLSimilarityIndex	0.286552	
591	13	NoOfLettersInURL	0.251936	
592	22	IsHTTPS	0.244181	

593	7	URLCharProb	0.184244
594	38	HasSocialNet	0.093251
595	45	HasCopyrightInfo	0.092814
596	15	NoOfDigitsInURL	0.087444
597	6	TLDLegitimateProb	0.064208
598	34	HasDescription	0.064188
599	14	LetterRatioInURL	0.041618
600	9	NoOfSubDomain	0.034921
601	39	HasSubmitButton	0.033295
602	29	HasFavicon	0.028812
603	23	LineOfCode	0.023595
604	32	NoOfURLRedirect	0.021891
605	5	CharContinuationRate	0.018712
606	<class 'numpy.ndarray'>		
607	Top 20 Features (Lasso):		
608		Feature	Importance
609	4	URLSimilarityIndex	0.237089
610	22	IsHTTPS	0.228563
611	45	HasCopyrightInfo	0.088124
612	38	HasSocialNet	0.076111
613	34	HasDescription	0.054220
614	9	NoOfSubDomain	0.034492
615	29	HasFavicon	0.007343
616	39	HasSubmitButton	0.004301
617	48	NoOfJS	0.002680
618	47	NoOfCSS	0.002125
619	28	URLTitleMatchScore	0.001026
620	46	NoOfImage	0.000836
621	50	NoOfEmptyRef	0.000452
622	49	NoOfSelfRef	0.000282
623	3	TLD	0.000169
624	51	NoOfExternalRef	0.000160
625	26	Title	0.000000
626	1	DomainLength	0.000000
627	2	IsDomainIP	0.000000
628	0	URLLength	0.000000
629	train_Ridge_Lasso.py	executed successfully.	
630			
631	Running train_decision_tree.py...		
632	URLLength DomainLength ... NoOfEmptyRef		
	NoOfExternalRef		

633	104874	-0.255920 28.0	-0.488527	...	1.0
634	176003	-0.304330 22.0	-0.707087	...	4.0
635	137445	-0.207510 0.0	-0.269966	...	0.0
636	38100	0.107152 11.0	1.150679	...	5.0
637	65088	-0.110691 135.0	0.167156	...	1.0
638
		...			
639	160997	-0.062281 8.0	0.385716	...	3.0
640	155379	-0.255920 0.0	-0.379246	...	0.0
641	174379	0.470225 2.0	1.806361	...	0.0
642	920	-0.183305 63.0	-0.160685	...	2.0
643	172904	-0.304330 0.0	-0.597807	...	0.0
644					
645	[188636 rows x 52 columns]				
646		URLLength	DomainLength	...	NoOfEmptyRef
		NoOfExternalRef			
647	173310	-0.231715 11.0	-0.379246	...	0.0
648	62486	-0.231715 127.0	-0.379246	...	0.0
649	14858	-0.255920 6.0	-0.488527	...	0.0
650	82943	-0.062281 208.0	0.385716	...	2.0
651	72539	-0.280125 9.0	-0.597807	...	0.0
652
		...			
653	109651	-0.207510 0.0	-0.379246	...	0.0
654	133536	-0.352739	-0.925648	...	1.0

```
654          45.0
655 192832   -0.086486      0.276436 ...      3.0
              19.0
656 92195    -0.134896      0.057875 ...      0.0
              17.0
657 153247   -0.255920      -0.488527 ...      0.0
              31.0
658
659 [47159 rows x 52 columns]
660 104874    1.0
661 176003    1.0
662 137445    0.0
663 38100     1.0
664 65088     1.0
665 ...
666 160997    1.0
667 155379    0.0
668 174379    0.0
669 920       1.0
670 172904    0.0
671 Name: label, Length: 188636, dtype: float64
672 173310    1.0
673 62486     1.0
674 14858     1.0
675 82943     1.0
676 72539     1.0
677 ...
678 109651    0.0
679 133536    1.0
680 192832    1.0
681 92195     1.0
682 153247    1.0
683 Name: label, Length: 47159, dtype: float64
684 Train set size: (188636, 52)
685 Test set size: (47159, 52)
686 Accuracy: 0.9789
687 Precision: 0.9887
688 Recall: 0.9742
689 F1-score: 0.9814
690 Mean Squared Error (MSE): 0.0211
691 Root Mean Squared Error (RMSE): 0.1454
```

```
692 R2 Score: 0.9136
693
694 Confusion Matrix:
695 [[19888 301]
696 [ 696 26274]]
697
698 Classification Report:
699
700
701          precision    recall   f1-score
702
703          support
704
705          accuracy           0.98
706
707          Feature  Importance
708
709 23          LineOfCode  0.901817
710 20  NoOfOtherSpecialCharsInURL  0.098183
711 0          URLLength  0.000000
712 1          DomainLength  0.000000
713 4          URLSimilarityIndex  0.000000
714 5          CharContinuationRate  0.000000
715 6          TLDLegitimateProb  0.000000
716 7          URLCharProb  0.000000
717 8          TLDLength  0.000000
718 9          NoOfSubDomain  0.000000
719 10         Has0bfuscation  0.000000
720 11         No0fbfuscatdChar  0.000000
721 12         ObfuscationRatio  0.000000
722 13         NoOfLettersInURL  0.000000
723 14         LetterRatioInURL  0.000000
724 15         NoOfDigitsInURL  0.000000
725 16         DigitRatioInURL  0.000000
726 17         NoOfEqualsInURL  0.000000
```

727	2		IsDomainIP	0.000000
728	3		TLD	0.000000
729	train_decision_tree.py	executed successfully.		
730				
731	Running train_knn.py...			
732		URLLength DomainLength ... NoOfEmptyRef		
		NoOfExternalRef		
733	104874	-0.255920 28.0	-0.488527 ...	1.0
734	176003	-0.304330 22.0	-0.707087 ...	4.0
735	137445	-0.207510 0.0	-0.269966 ...	0.0
736	38100	0.107152 11.0	1.150679 ...	5.0
737	65088	-0.110691 135.0	0.167156 ...	1.0
738
		...		
739	160997	-0.062281 8.0	0.385716 ...	3.0
740	155379	-0.255920 0.0	-0.379246 ...	0.0
741	174379	0.470225 2.0	1.806361 ...	0.0
742	920	-0.183305 63.0	-0.160685 ...	2.0
743	172904	-0.304330 0.0	-0.597807 ...	0.0
744				
745	[188636 rows x 52 columns]			
746		URLLength DomainLength ... NoOfEmptyRef		
		NoOfExternalRef		
747	173310	-0.231715 11.0	-0.379246 ...	0.0
748	62486	-0.231715 127.0	-0.379246 ...	0.0
749	14858	-0.255920 6.0	-0.488527 ...	0.0
750	82943	-0.062281 208.0	0.385716 ...	2.0

751	72539	-0.280125 9.0	-0.597807	...	0.0
752
753	109651	...	-0.379246	...	0.0
		0.0			
754	133536	-0.352739 45.0	-0.925648	...	1.0
755	192832	-0.086486 19.0	0.276436	...	3.0
756	92195	-0.134896 17.0	0.057875	...	0.0
757	153247	-0.255920 31.0	-0.488527	...	0.0
758					
759		[47159 rows x 52 columns]			
760	104874	1.0			
761	176003	1.0			
762	137445	0.0			
763	38100	1.0			
764	65088	1.0			
765		...			
766	160997	1.0			
767	155379	0.0			
768	174379	0.0			
769	920	1.0			
770	172904	0.0			
771		Name: label, Length: 188636, dtype: float64			
772	173310	1.0			
773	62486	1.0			
774	14858	1.0			
775	82943	1.0			
776	72539	1.0			
777		...			
778	109651	0.0			
779	133536	1.0			
780	192832	1.0			
781	92195	1.0			
782	153247	1.0			
783		Name: label, Length: 47159, dtype: float64			
784		Train set size: (188636, 52)			

```
785 Test set size: (47159, 52)
786 Accuracy: 0.9262
787 Precision: 0.9430
788 Recall: 0.9271
789 F1-score: 0.9350
790
791 Confusion Matrix:
792 [[18677 1512]
793 [ 1966 25004]]
794
795 Classification Report:
796              precision    recall   f1-score
797      support
798
799          0.0        0.90      0.93      0.91
800      20189
801          1.0        0.94      0.93      0.93
802      26970
803
804      accuracy
805      47159
806          0.93        0.92      0.93      0.92
807      47159
808      weighted avg
809      47159
810
811      accuracy
812      47159
813
814      macro avg
815      47159
816
817      weighted avg
818      47159
819
820      accuracy
821      47159
822
823      macro avg
824      47159
825
826      weighted avg
827      47159
828
829      accuracy
830      47159
831
832      macro avg
833      47159
834
835      weighted avg
836      47159
837
838      accuracy
839      47159
840
841      macro avg
842      47159
843
844      weighted avg
845      47159
846
847      accuracy
848      47159
849
850      macro avg
851      47159
852
853      weighted avg
854      47159
855
856      accuracy
857      47159
858
859      macro avg
860      47159
861
862      weighted avg
863      47159
864
865      accuracy
866      47159
867
868      macro avg
869      47159
870
871      weighted avg
872      47159
873
874      accuracy
875      47159
876
877      macro avg
878      47159
879
880      weighted avg
881      47159
882
883      accuracy
884      47159
885
886      macro avg
887      47159
888
889      weighted avg
890      47159
891
892      accuracy
893      47159
894
895      macro avg
896      47159
897
898      weighted avg
899      47159
900
901      accuracy
902      47159
903
904      macro avg
905      47159
906
907      weighted avg
908      47159
909
910      accuracy
911      47159
912
913      macro avg
914      47159
915
916      weighted avg
917      47159
918
919      accuracy
920      47159
921
922      macro avg
923      47159
924
925      weighted avg
926      47159
927
928      accuracy
929      47159
930
931      macro avg
932      47159
933
934      weighted avg
935      47159
936
937      accuracy
938      47159
939
940      macro avg
941      47159
942
943      weighted avg
944      47159
945
946      accuracy
947      47159
948
949      macro avg
950      47159
951
952      weighted avg
953      47159
954
955      accuracy
956      47159
957
958      macro avg
959      47159
960
961      weighted avg
962      47159
963
964      accuracy
965      47159
966
967      macro avg
968      47159
969
970      weighted avg
971      47159
972
973      accuracy
974      47159
975
976      macro avg
977      47159
978
979      weighted avg
980      47159
981
982      accuracy
983      47159
984
985      macro avg
986      47159
987
988      weighted avg
989      47159
990
991      accuracy
992      47159
993
994      macro avg
995      47159
996
997      weighted avg
998      47159
999
1000      accuracy
1001      47159
1002
1003      macro avg
1004      47159
1005
1006      weighted avg
1007      47159
1008
1009      accuracy
1010      47159
1011
1012      macro avg
1013      47159
1014
1015      weighted avg
1016      47159
1017
1018      accuracy
1019      47159
1020
1021      macro avg
1022      47159
1023
1024      weighted avg
1025      47159
1026
1027      accuracy
1028      47159
1029
1030      macro avg
1031      47159
1032
1033      weighted avg
1034      47159
1035
1036      accuracy
1037      47159
1038
1039      macro avg
1040      47159
1041
1042      weighted avg
1043      47159
1044
1045      accuracy
1046      47159
1047
1048      macro avg
1049      47159
1050
1051      weighted avg
1052      47159
1053
1054      accuracy
1055      47159
1056
1057      macro avg
1058      47159
1059
1060      weighted avg
1061      47159
1062
1063      accuracy
1064      47159
1065
1066      macro avg
1067      47159
1068
1069      weighted avg
1070      47159
1071
1072      accuracy
1073      47159
1074
1075      macro avg
1076      47159
1077
1078      weighted avg
1079      47159
1080
1081      accuracy
1082      47159
1083
1084      macro avg
1085      47159
1086
1087      weighted avg
1088      47159
1089
1090      accuracy
1091      47159
1092
1093      macro avg
1094      47159
1095
1096      weighted avg
1097      47159
1098
1099      accuracy
1100      47159
1101
1102      macro avg
1103      47159
1104
1105      weighted avg
1106      47159
1107
1108      accuracy
1109      47159
1110
1111      macro avg
1112      47159
1113
1114      weighted avg
1115      47159
1116
1117      accuracy
1118      47159
1119
1120      macro avg
1121      47159
1122
1123      weighted avg
1124      47159
1125
1126      accuracy
1127      47159
1128
1129      macro avg
1130      47159
1131
1132      weighted avg
1133      47159
1134
1135      accuracy
1136      47159
1137
1138      macro avg
1139      47159
1140
1141      weighted avg
1142      47159
1143
1144      accuracy
1145      47159
1146
1147      macro avg
1148      47159
1149
1150      weighted avg
1151      47159
1152
1153      accuracy
1154      47159
1155
1156      macro avg
1157      47159
1158
1159      weighted avg
1160      47159
1161
1162      accuracy
1163      47159
1164
1165      macro avg
1166      47159
1167
1168      weighted avg
1169      47159
1170
1171      accuracy
1172      47159
1173
1174      macro avg
1175      47159
1176
1177      weighted avg
1178      47159
1179
1180      accuracy
1181      47159
1182
1183      macro avg
1184      47159
1185
1186      weighted avg
1187      47159
1188
1189      accuracy
1190      47159
1191
1192      macro avg
1193      47159
1194
1195      weighted avg
1196      47159
1197
1198      accuracy
1199      47159
1200
1201      macro avg
1202      47159
1203
1204      weighted avg
1205      47159
1206
1207      accuracy
1208      47159
1209
1210      macro avg
1211      47159
1212
1213      weighted avg
1214      47159
1215
1216      accuracy
1217      47159
1218
1219      macro avg
1220      47159
1221
1222      weighted avg
1223      47159
1224
1225      accuracy
1226      47159
1227
1228      macro avg
1229      47159
1230
1231      weighted avg
1232      47159
1233
1234      accuracy
1235      47159
1236
1237      macro avg
1238      47159
1239
1240      weighted avg
1241      47159
1242
1243      accuracy
1244      47159
1245
1246      macro avg
1247      47159
1248
1249      weighted avg
1250      47159
1251
1252      accuracy
1253      47159
1254
1255      macro avg
1256      47159
1257
1258      weighted avg
1259      47159
1260
1261      accuracy
1262      47159
1263
1264      macro avg
1265      47159
1266
1267      weighted avg
1268      47159
1269
1270      accuracy
1271      47159
1272
1273      macro avg
1274      47159
1275
1276      weighted avg
1277      47159
1278
1279      accuracy
1280      47159
1281
1282      macro avg
1283      47159
1284
1285      weighted avg
1286      47159
1287
1288      accuracy
1289      47159
1290
1291      macro avg
1292      47159
1293
1294      weighted avg
1295      47159
1296
1297      accuracy
1298      47159
1299
1300      macro avg
1301      47159
1302
1303      weighted avg
1304      47159
1305
1306      accuracy
1307      47159
1308
1309      macro avg
1310      47159
1311
1312      weighted avg
1313      47159
1314
1315      accuracy
1316      47159
1317
1318      macro avg
1319      47159
1320
1321      weighted avg
1322      47159
1323
1324      accuracy
1325      47159
1326
1327      macro avg
1328      47159
1329
1330      weighted avg
1331      47159
1332
1333      accuracy
1334      47159
1335
1336      macro avg
1337      47159
1338
1339      weighted avg
1340      47159
1341
1342      accuracy
1343      47159
1344
1345      macro avg
1346      47159
1347
1348      weighted avg
1349      47159
1350
1351      accuracy
1352      47159
1353
1354      macro avg
1355      47159
1356
1357      weighted avg
1358      47159
1359
1360      accuracy
1361      47159
1362
1363      macro avg
1364      47159
1365
1366      weighted avg
1367      47159
1368
1369      accuracy
1370      47159
1371
1372      macro avg
1373      47159
1374
1375      weighted avg
1376      47159
1377
1378      accuracy
1379      47159
1380
1381      macro avg
1382      47159
1383
1384      weighted avg
1385      47159
1386
1387      accuracy
1388      47159
1389
1390      macro avg
1391      47159
1392
1393      weighted avg
1394      47159
1395
1396      accuracy
1397      47159
1398
1399      macro avg
1400      47159
1401
1402      weighted avg
1403      47159
1404
1405      accuracy
1406      47159
1407
1408      macro avg
1409      47159
1410
1411      weighted avg
1412      47159
1413
1414      accuracy
1415      47159
1416
1417      macro avg
1418      47159
1419
1420      weighted avg
1421      47159
1422
1423      accuracy
1424      47159
1425
1426      macro avg
1427      47159
1428
1429      weighted avg
1430      47159
1431
1432      accuracy
1433      47159
1434
1435      macro avg
1436      47159
1437
1438      weighted avg
1439      47159
1440
1441      accuracy
1442      47159
1443
1444      macro avg
1445      47159
1446
1447      weighted avg
1448      47159
1449
1450      accuracy
1451      47159
1452
1453      macro avg
1454      47159
1455
1456      weighted avg
1457      47159
1458
1459      accuracy
1460      47159
1461
1462      macro avg
1463      47159
1464
1465      weighted avg
1466      47159
1467
1468      accuracy
1469      47159
1470
1471      macro avg
1472      47159
1473
1474      weighted avg
1475      47159
1476
1477      accuracy
1478      47159
1479
1480      macro avg
1481      47159
1482
1483      weighted avg
1484      47159
1485
1486      accuracy
1487      47159
1488
1489      macro avg
1490      47159
1491
1492      weighted avg
1493      47159
1494
1495      accuracy
1496      47159
1497
1498      macro avg
1499      47159
1500
1501      weighted avg
1502      47159
1503
1504      accuracy
1505      47159
1506
1507      macro avg
1508      47159
1509
1510      weighted avg
1511      47159
1512
1513      accuracy
1514      47159
1515
1516      macro avg
1517      47159
1518
1519      weighted avg
1520      47159
1521
1522      accuracy
1523      47159
1524
1525      macro avg
1526      47159
1527
1528      weighted avg
1529      47159
1530
1531      accuracy
1532      47159
1533
1534      macro avg
1535      47159
1536
1537      weighted avg
1538      47159
1539
1540      accuracy
1541      47159
1542
1543      macro avg
1544      47159
1545
1546      weighted avg
1547      47159
1548
1549      accuracy
1550      47159
1551
1552      macro avg
1553      47159
1554
1555      weighted avg
1556      47159
1557
1558      accuracy
1559      47159
1560
1561      macro avg
1562      47159
1563
1564      weighted avg
1565      47159
1566
1567      accuracy
1568      47159
1569
1570      macro avg
1571      47159
1572
1573      weighted avg
1574      47159
1575
1576      accuracy
1577      47159
1578
1579      macro avg
1580      47159
1581
1582      weighted avg
1583      47159
1584
1585      accuracy
1586      47159
1587
1588      macro avg
1589      47159
1590
1591      weighted avg
1592      47159
1593
1594      accuracy
1595      47159
1596
1597      macro avg
1598      47159
1599
15100      weighted avg
15101      47159
15102
15103      accuracy
15104      47159
15105
15106      macro avg
15107      47159
15108
15109      weighted avg
15110      47159
15111
15112      accuracy
15113      47159
15114
15115      macro avg
15116      47159
15117
15118      weighted avg
15119      47159
15120
15121      accuracy
15122      47159
15123
15124      macro avg
15125      47159
15126
15127      weighted avg
15128      47159
15129
15130      accuracy
15131      47159
15132
15133      macro avg
15134      47159
15135
15136      weighted avg
15137      47159
15138
15139      accuracy
15140      47159
15141
15142      macro avg
15143      47159
15144
15145      weighted avg
15146      47159
15147
15148      accuracy
15149      47159
15150
15151      macro avg
15152      47159
15153
15154      weighted avg
15155      47159
15156
15157      accuracy
15158      47159
15159
15160      macro avg
15161      47159
15162
15163      weighted avg
15164      47159
15165
15166      accuracy
15167      47159
15168
15169      macro avg
15170      47159
15171
15172      weighted avg
15173      47159
15174
15175      accuracy
15176      47159
15177
15178      macro avg
15179      47159
15180
15181      weighted avg
15182      47159
15183
15184      accuracy
15185      47159
15186
15187      macro avg
15188      47159
15189
15190      weighted avg
15191      47159
15192
15193      accuracy
15194      47159
15195
15196      macro avg
15197      47159
15198
15199      weighted avg
15200      47159
15201
15202      accuracy
15203      47159
15204
15205      macro avg
15206      47159
15207
15208      weighted avg
15209      47159
15210
15211      accuracy
15212      47159
15213
15214      macro avg
15215      47159
15216
15217      weighted avg
15218      47159
15219
15220      accuracy
15221      47159
15222
15223      macro avg
15224      47159
15225
15226      weighted avg
15227      47159
15228
15229      accuracy
15230      47159
15231
15232      macro avg
15233      47159
15234
15235      weighted avg
15236      47159
15237
15238      accuracy
15239      47159
15240
15241      macro avg
15242      47159
15243
15244      weighted avg
15245      47159
15246
15247      accuracy
15248      47159
15249
15250      macro avg
15251      47159
15252
15253      weighted avg
15254      47159
15255
15256      accuracy
15257      47159
15258
15259      macro avg
15260      47159
15261
15262      weighted avg
15263      47159
15264
15265      accuracy
15266      47159
15267
15268      macro avg
15269      47159
15270
15271      weighted avg
15272      47159
15273
15274      accuracy
15275      47159
15276
15277      macro avg
15278      47159
15279
15280      weighted avg
15281      47159
15282
15283      accuracy
15284      47159
15285
15286      macro avg
15287      47159
15288
15289      weighted avg
15290      47159
15291
15292      accuracy
15293      47159
15294
15295      macro avg
15296      47159
15297
15298      weighted avg
15299      47159
15300
15301      accuracy
15302      47159
15303
15304      macro avg
15305      47159
15306
15307      weighted avg
15308      47159
15309
15310      accuracy
15311      47159
15312
15313      macro avg
15314      47159
15315
15316      weighted avg
15317      47159
15318
15319      accuracy
15320      47159
15321
15322      macro avg
15323      47159
15324
15325      weighted avg
15326      47159
15327
15328      accuracy
15329      47159
15330
15331      macro avg
15332      47159
15333
15334      weighted avg
15335      47159
15336
15337      accuracy
15338      47159
15339
15340      macro avg
15341      47159
15342
15343      weighted avg
15344      47159
15345
15346      accuracy
15347      47159
15348
15349      macro avg
15350      47159
15351
15352      weighted avg
15353      47159
15354
15355      accuracy
15356      47159
15357
15358      macro avg
15359      47159
15360
15361      weighted avg
15362      47159
15363
15364      accuracy
15365      47159
15366
15367      macro avg
15368      47159
15369
15370      weighted avg
15371      47159
15372
15373      accuracy
15374      47159
15375
15376      macro avg
15377      47159
15378
15379      weighted avg
15380      47159
15381
15382      accuracy
15383      47159
15384
15385      macro avg
15386      47159
15387
15388      weighted avg
15389      47159
15390
15391      accuracy
15392      47159
15393
15394      macro avg
15395      47159
15396
15397      weighted avg
15398      47159
15399
15400      accuracy
15401      47159
15402
15403      macro avg
15404      47159
15405
15406      weighted avg
15407      47159
15408
15409      accuracy
15410      47159
15411
15412      macro avg
15413      47159
15414
15415      weighted avg
15416      47159
15417
15418      accuracy
15419      47159
15420
15421      macro avg
15422      47159
15423
15424      weighted avg
15425      47159
15426
15427      accuracy
15428      47159
15429
15430      macro avg
15431      47159
15432
15433      weighted avg
15434      47159
15435
15436      accuracy
15437      47159
15438
15439      macro avg
15440      47159
15441
15442      weighted avg
15443      47159
15444
15445      accuracy
15446      47159
15447
15448      macro avg
15449      47159
15450
15451      weighted avg
15452      47159
15453
15454      accuracy
15455      47159
15456
15457      macro avg
15458      47159
15459
15460      weighted avg
15461      47159
15462
15463      accuracy
15464      47159
15465
15466      macro avg
15467      47159
15468
15469      weighted avg
15470      47159
15471
15472      accuracy
15473      47159
15474
15475      macro avg
15476      47159
15477
15478      weighted avg
15479      47159
15480
15481      accuracy
15482      47159
15483
15484      macro avg
15485      47159
15486
15487      weighted avg
15488      47159
15489
15490      accuracy
15491      47159
15492
15493      macro avg
15494      47159
15495
15496      weighted avg
15497      47159
15498
15499      accuracy
15500      47159
15501
15502      macro avg
15503      47159
15504
15505      weighted avg
15506      47159
15507
15508      accuracy
15509      47159
15510
15511      macro avg
15512      47159
15513
15514      weighted avg
15515      47159
15516
15517      accuracy
15518      47159
15519
15520      macro avg
15521      47159
15522
15523      weighted avg
15524      47159
15525
15526      accuracy
15527      47159
15528
15529      macro avg
15530      47159
15531
15532      weighted avg
15533      47159
15534
15535      accuracy
15536      47159
15537
15538      macro avg
15539      47159
15540
15541      weighted avg
15542      47159
15543
15544      accuracy
15545      47159
15546
15547      macro avg
15548      47159
15549
15550      weighted avg
15551      47159
15552
15553      accuracy
15554      47159
15555
15556      macro avg
15557      47159
15558
15559      weighted avg
15560      47159
15561
15562      accuracy
15563      47159
15564
15565      macro avg
15566      47159
15567
15568      weighted avg
15569      47159
15570
15571      accuracy
15572      47159
15573
15574      macro avg
15575      47159
15576
15577      weighted avg
15578      47159
15579
15580      accuracy
15581      47159
15582
15583      macro avg
15584      47159
15585
15586      weighted avg
15587      47159
15588
15589      accuracy
155
```

```

812 11444000e+04
813 7.80460000e+03 0.00000000e+00 1.28856000e+04 4.
   68580000e+03
814 6.11660000e+04 0.00000000e+00 9.80760000e+03 1.
   28900000e+04
815 1.41156000e+04 6.99140000e+03 7.96060000e+03 1.
   27468000e+04
816 0.00000000e+00 1.11816000e+04 5.31389400e+05 2.
   06925400e+05
817 2.95257200e+05 7.18720800e+05 7.94442000e+04 6.
   18172400e+05]
818 Top 20 Features (KNN - Distance-Based Importance):
819          Feature  Importance
820 26          Title    19.864282
821 49        NoOfSelfRef 15.240374
822 51        NoOfExternalRef 13.108259
823 46        NoOfImage   11.268038
824 3          TLD      10.157751
825 48        NoOfJS     6.260888
826 47        NoOfCSS    4.387824
827 28        URLTitleMatchScore 3.768925
828 50        NoOfEmptyRef 1.684603
829 36        NoOfFrame   1.297016
830 14        LetterRatioInURL 0.614652
831 1          DomainLength 0.443412
832 29        HasFavicon  0.317789
833 40        HasHiddenFields 0.299319
834 39        HasSubmitButton 0.273331
835 34        HasDescription 0.273237
836 43          Pay      0.270294
837 30        Robots    0.259654
838 4        URLSimilarityIndex 0.248168
839 45        HasCopyrightInfo 0.237104
840 train_knn.py executed successfully.
841
842 Running train_random_forest.py...
843          URLLength  DomainLength ...  NoOfEmptyRef
   NoOfExternalRef
844 104874  -0.255920     -0.488527 ...  1.0
   28.0
845 176003  -0.304330     -0.707087 ...  4.0

```

845		22.0				
846	137445	-0.207510 0.0	-0.269966	...		0.0
847	38100	0.107152 11.0	1.150679	...		5.0
848	65088	-0.110691 135.0	0.167156	...		1.0
849
		...				
850	160997	-0.062281 8.0	0.385716	...		3.0
851	155379	-0.255920 0.0	-0.379246	...		0.0
852	174379	0.470225 2.0	1.806361	...		0.0
853	920	-0.183305 63.0	-0.160685	...		2.0
854	172904	-0.304330 0.0	-0.597807	...		0.0
855						
856	[188636 rows x 52 columns]					
857		URLLength	DomainLength	...	NoOfEmptyRef	
		NoOfExternalRef				
858	173310	-0.231715 11.0	-0.379246	...		0.0
859	62486	-0.231715 127.0	-0.379246	...		0.0
860	14858	-0.255920 6.0	-0.488527	...		0.0
861	82943	-0.062281 208.0	0.385716	...		2.0
862	72539	-0.280125 9.0	-0.597807	...		0.0
863
		...				
864	109651	-0.207510 0.0	-0.379246	...		0.0
865	133536	-0.352739 45.0	-0.925648	...		1.0
866	192832	-0.086486 19.0	0.276436	...		3.0

```
867 92195 -0.134896 0.057875 ... 0.0
     17.0
868 153247 -0.255920 -0.488527 ... 0.0
     31.0
869
870 [47159 rows x 52 columns]
871 104874 1.0
872 176003 1.0
873 137445 0.0
874 38100 1.0
875 65088 1.0
876 ...
877 160997 1.0
878 155379 0.0
879 174379 0.0
880 920 1.0
881 172904 0.0
882 Name: label, Length: 188636, dtype: float64
883 173310 1.0
884 62486 1.0
885 14858 1.0
886 82943 1.0
887 72539 1.0
888 ...
889 109651 0.0
890 133536 1.0
891 192832 1.0
892 92195 1.0
893 153247 1.0
894 Name: label, Length: 47159, dtype: float64
895 Train set size: (188636, 52)
896 Test set size: (47159, 52)
897 Random Forest Accuracy: 1.0000
898 Random Forest Precision: 1.0000
899 Random Forest Recall: 1.0000
900 Random Forest F1 Score: 1.0000
901 Confusion Matrix:
902 [[20189 0]
903 [ 0 26970]]
904
905 Classification Report:
```

		precision	recall	f1-score
906	support			
907				
908	0.0	1.00	1.00	1.00
909	20189	1.0	1.00	1.00
910	26970			
911	accuracy			1.00
912	47159			
913	macro avg	1.00	1.00	1.00
914	47159			
915	weighted avg	1.00	1.00	1.00
916	47159			
917	Top 20 Features (Random Forest):			
918		Feature	Importance	
919	4	URLSimilarityIndex	0.213967	
920	23	LineOfCode	0.139181	
921	51	NoOfExternalRef	0.115639	
922	46	NoOfImage	0.076205	
923	49	NoOfSelfRef	0.072740	
924	48	NoOfJS	0.071929	
925	47	NoOfCSS	0.052851	
926	45	HasCopyrightInfo	0.052139	
927	34	HasDescription	0.035773	
928	38	HasSocialNet	0.025142	
929	20	NoOfOtherSpecialCharsInURL	0.023650	
930	22	IsHTTPS	0.019496	
931	24	LargestLineLength	0.017540	
932	27	DomainTitleMatchScore	0.011389	
933	15	NoOfDigitsInURL	0.010145	
934	0	URLLength	0.009391	
935	28	URLTitleMatchScore	0.007709	
936	16	DigitRatioInURL	0.005370	
937	21	SpacialCharRatioInURL	0.005239	
938	39	HasSubmitButton	0.004912	
939	Top 20 features saved to 'top_20_features.json'			
940	Random Forest model saved successfully!			
941	train_random_forest.py executed successfully.			

941	Running	train_svc.py...				
942		URLLength	DomainLength	...	NoOfEmptyRef	
		NoOfExternalRef				
943	104874	-0.255920 28.0	-0.488527	...		1.0
944	176003	-0.304330 22.0	-0.707087	...		4.0
945	137445	-0.207510 0.0	-0.269966	...		0.0
946	38100	0.107152 11.0	1.150679	...		5.0
947	65088	-0.110691 135.0	0.167156	...		1.0
948
		...				
949	160997	-0.062281 8.0	0.385716	...		3.0
950	155379	-0.255920 0.0	-0.379246	...		0.0
951	174379	0.470225 2.0	1.806361	...		0.0
952	920	-0.183305 63.0	-0.160685	...		2.0
953	172904	-0.304330 0.0	-0.597807	...		0.0
954						
955	[188636 rows x 52 columns]					
956		URLLength	DomainLength	...	NoOfEmptyRef	
		NoOfExternalRef				
957	173310	-0.231715 11.0	-0.379246	...		0.0
958	62486	-0.231715 127.0	-0.379246	...		0.0
959	14858	-0.255920 6.0	-0.488527	...		0.0
960	82943	-0.062281 208.0	0.385716	...		2.0
961	72539	-0.280125 9.0	-0.597807	...		0.0
962
		...				

```
963 109651 -0.207510 -0.379246 ... 0.0
      0.0
964 133536 -0.352739 -0.925648 ... 1.0
      45.0
965 192832 -0.086486 0.276436 ... 3.0
      19.0
966 92195 -0.134896 0.057875 ... 0.0
      17.0
967 153247 -0.255920 -0.488527 ... 0.0
      31.0
968
969 [47159 rows x 52 columns]
970 104874 1.0
971 176003 1.0
972 137445 0.0
973 38100 1.0
974 65088 1.0
975 ...
976 160997 1.0
977 155379 0.0
978 174379 0.0
979 920 1.0
980 172904 0.0
981 Name: label, Length: 188636, dtype: float64
982 173310 1.0
983 62486 1.0
984 14858 1.0
985 82943 1.0
986 72539 1.0
987 ...
988 109651 0.0
989 133536 1.0
990 192832 1.0
991 92195 1.0
992 153247 1.0
993 Name: label, Length: 47159, dtype: float64
994 Train set size: (188636, 52)
995 Test set size: (47159, 52)
996 SVM Accuracy: 0.9820
997 SVM Precision: 0.9930
998 SVM Recall: 0.9754
```

```

999 SVM F1 Score: 0.9841
1000 Confusion Matrix:
1001 [[20004 185]
1002 [ 663 26307]]
1003
1004 Classification Report:
1005              precision    recall   f1-score
1006             support
1007
1008             0.0        0.97      0.99      0.98
1009             20189
1010             1.0        0.99      0.98      0.98
1011             26970
1012
1013             accuracy          0.98
1014             47159
1015             macro avg      0.98      0.98      0.98
1016             47159
1017             weighted avg  0.98      0.98      0.98
1018             47159
1019
1020             Top 20 Features (SVC):
1021             Feature      Importance
1022             46           27244.217965
1023             51           26804.149666
1024             49           26328.209831
1025             48           13417.562197
1026             47           10393.804512
1027             4            URLSimilarityIndex 7069.872459
1028             28           URLTitleMatchScore 6775.690849
1029             9            NoOfSubDomain     3248.073374
1030             50           NoOfEmptyRef     3071.000000
1031             14           LetterRatioInURL 2868.199640
1032             34           HasDescription    2595.327311
1033             45           HasCopyrightInfo 2552.512221
1034             36           NoOfiFrame       2410.000000
1035             22           IsHTTPS         2314.327311
1036             38           HasSocialNet    2057.523894
1037             20           NoOfOtherSpecialCharsInURL 1979.310621
1038             27           DomainTitleMatchScore 1587.803801
1039             13           NoOfLettersInURL 1323.152521

```

		URLLength	1085.591257
		HasSubmitButton	1026.000000
1034	0	URLLength	1085.591257
1035	39	HasSubmitButton	1026.000000
1036 train_svc.py executed successfully.			
1037			
1038 Running model_Evaluation.py...			
1039 Sorted Model Performance (by F1-Score):			
1040			
1041		Accuracy	Precision
	F1-Score		Recall
1042	Random Forest	1.000000	1.000000
	1.000000		
1043	Ridge Regression	0.999406	0.998963
	0.999481		1.000000
1044	Lasso Regression	0.999109	0.998445
	0.999222		1.000000
1045	Logistic Regression	0.992812	0.995756
	0.993702		0.991657
1046	SVC	0.982018	0.993017
	0.984138		0.975417
1047	Decision Tree	0.978859	0.988674
	0.981380		0.974194
1048	KNN	0.926249	0.942978
	0.934974		0.927104
1049	model_Evaluation.py	executed successfully.	
1050			
1051	Running random_forest_top20_vs_all.py...		
1052	Performance Comparison:		
1053		Random Forest (All Features)	Random
	Forest (Top 20 Features)		
1054	Accuracy	0.999979	
		1.0	
1055	Precision	0.999963	
		1.0	
1056	Recall	1.000000	
		1.0	
1057	F1 Score	0.999982	
		1.0	
1058	random_forest_top20_vs_all.py	executed successfully	
	.		
1059			
1060	Running model_testing.py...		

```
1061 Sample Inputs:  
1062     URLLength DomainLength ... NoOfEmptyRef  
      NoOfExternalRef  
1063 0 -0.086486 0.276436 ... 0.0  
      124.0  
1064 1 -0.280125 -0.597807 ... 0.0  
      208.0  
1065 2 -0.134896 0.057875 ... 2.0  
      5.0  
1066 3 -0.207510 -0.269966 ... 1.0  
      31.0  
1067 4 -0.038077 0.494997 ... 1.0  
      85.0  
1068 5 -0.110691 0.167156 ... 0.0  
      14.0  
1069 6 -0.231715 -0.379246 ... 2.0  
      17.0  
1070 7 -0.231715 -0.379246 ... 0.0  
      15.0  
1071 8 -0.134896 0.057875 ... 1.0  
      208.0  
1072 9 -0.376944 -1.034928 ... 6.0  
      65.0  
1073  
1074 [10 rows x 52 columns]  
1075  
1076 Predictions for Sample Inputs:  
1077 [1. 1. 1. 1. 1. 1. 1. 1. 1.]  
1078  
1079 Train Accuracy: 1.0000  
1080 Test Accuracy: 1.0000  
1081 model_testing.py executed successfully.  
1082  
1083 Running app.py...  
1084 * Serving Flask app 'app'  
1085 * Debug mode: on  
1086 WARNING: This is a development server. Do not use  
      it in a production deployment. Use a production  
      WSGI server instead.  
1087 * Running on http://127.0.0.1:5000  
1088 Press CTRL+C to quit
```

```
1089 * Restarting with stat
1090 * Debugger is active!
1091 * Debugger PIN: 265-784-434
1092 127.0.0.1 - - [25/Feb/2025 22:08:20] "GET / HTTP/1.
1" 200 -
1093 127.0.0.1 - - [25/Feb/2025 22:08:20] "GET /static/
    styles.css HTTP/1.1" 404 -
1094 127.0.0.1 - - [25/Feb/2025 22:08:20] "GET /static/
    images/p1.jpg HTTP/1.1" 304 -
1095 127.0.0.1 - - [25/Feb/2025 22:08:20] "GET /favicon.
    ico HTTP/1.1" 404 -
1096 Feature values found. Making prediction...
1097 Prediction result: 0.0
1098 Legitimate. You are Safe !!!
1099 Returning prediction result to frontend.
1100 127.0.0.1 - - [25/Feb/2025 22:08:37] "POST /predict
    HTTP/1.1" 200 -
1101 Page loaded successfully
1102 Jaccard Calculation Started
1103 Jaccard Calculation Completed
1104 Features extracted successfully from live site.
1105 Feature values found. Making prediction...
1106 127.0.0.1 - - [25/Feb/2025 22:09:39] "POST /predict
    HTTP/1.1" 200 -
1107 Prediction result: 1.0
1108 Phishing!! Be Careful !!!
1109 Returning prediction result to frontend.
1110 Feature values found. Making prediction...
1111 Prediction result: 0.0
1112 Legitimate. You are Safe !!!
1113 Returning prediction result to frontend.
1114 127.0.0.1 - - [25/Feb/2025 22:10:06] "POST /predict
    HTTP/1.1" 200 -
1115 Page loaded successfully
1116 Jaccard Calculation Started
1117 Jaccard Calculation Completed
1118 Features extracted successfully from live site.
1119 Feature values found. Making prediction...
1120 Prediction result: 1.0
1121 Phishing!! Be Careful !!!
1122 Returning prediction result to frontend.
```

```
1123 127.0.0.1 - - [25/Feb/2025 22:11:40] "POST /predict  
HTTP/1.1" 200 -  
1124
```