```python
# Importing dataset
import pandas as pd
import numpy as np
import seaborn as sns
```

```python
data = pd.read_csv("dataset_website.csv")
data
```

	index	having_IPhaving_IP_Address	URL_URL_Length	Shortining_Service	having_At_Symbol	double_slash_redirecting	Prefix_Suffix	having_Sub_Domain	SSLfinal_State	Domain_registration_length	...	popUpWindow	Iframe	age_of_domain	DNSRecord	web_traffic	Page_Rank	Google_Index	Links_pointing_to_page	Statistical_report	Result
0	1	-1	1	1	1	-1	-1	-1	-1	-1	...	1	1	-1	-1	-1	-1	1	1	-1	-1
1	2	1	1	1	1	1	-1	0	1	-1	...	1	1	-1	-1	0	-1	1	1	1	-1
2	3	1	0	1	1	1	-1	-1	-1	-1	...	1	1	1	-1	1	-1	1	0	-1	-1
3	4	1	0	1	1	1	-1	-1	-1	1	...	1	1	-1	-1	1	-1	1	-1	1	-1
4	5	1	0	-1	1	1	-1	1	1	-1	...	-1	1	-1	-1	0	-1	1	1	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
11050	11050	1	-1	1	-1	1	1	1	1	-1	...	-1	-1	1	1	-1	-1	1	1	1	1

index		having_IPhaving_IP_Address	URLURL_Length	Shortining_Service	having_At_Symbol	double_slash_redirecting	Prefix_Suffix	having_Sub_Domain	SSLfinal_State	Domain_registration_length	...	popUpWidnow	Iframe	age_of_domain	DNSRecord	web_traffic	Page_Rank	Google_Index	Links_pointing_to_page	Statistical_report	Result
1 1 0 5 1	1 1 0 5 2	-1	1	1	-1	-1	-1	1	-1	-1	...	-1	1	1	1	1	1	1	-1	1	-1
1 1 0 5 2	1 1 0 5 3	1	-1	1	1	1	-1	1	-1	-1	...	1	1	1	1	1	-1	1	0	1	-1
1 1 0 5 3	1 1 0 5 4	-1	-1	1	1	1	-1	-1	-1	1	...	-1	1	1	1	1	-1	1	1	1	-1
1 1 0 5 4	1 1 0 5 5	-1	-1	1	1	1	-1	-1	-1	1	...	1	1	-1	1	-1	-1	-1	1	-1	-1

11055 rows × 32 columns

In [4]:

```
data.head()
data.shape
```

Out[4]:

```
(11055, 32)
```

In [5]:

```
data.size
data.info()
```

```
RangeIndex: 11055 entries, 0 to 11054
Data columns (total 32 columns):
 #    Column                        Non-Null Count   Dtype
---   ------                        --------------   -----
 0    index                         11055 non-null   int64
 1    having_IPhaving_IP_Address    11055 non-null   int64
 2    URLURL_Length                 11055 non-null   int64
 3    Shortining_Service            11055 non-null   int64
 4    having_At_Symbol              11055 non-null   int64
```

```
5   double_slash_redirecting      11055 non-null   int64
6   Prefix_Suffix                 11055 non-null   int64
7   having_Sub_Domain             11055 non-null   int64
8   SSLfinal_State                11055 non-null   int64
9   Domain_registeration_length   11055 non-null   int64
10  Favicon                       11055 non-null   int64
11  port                          11055 non-null   int64
12  HTTPS_token                   11055 non-null   int64
13  Request_URL                   11055 non-null   int64
14  URL_of_Anchor                 11055 non-null   int64
15  Links_in_tags                 11055 non-null   int64
16  SFH                           11055 non-null   int64
17  Submitting_to_email           11055 non-null   int64
18  Abnormal_URL                  11055 non-null   int64
19  Redirect                      11055 non-null   int64
20  on_mouseover                  11055 non-null   int64
21  RightClick                    11055 non-null   int64
22  popUpWidnow                   11055 non-null   int64
23  Iframe                        11055 non-null   int64
24  age_of_domain                 11055 non-null   int64
25  DNSRecord                     11055 non-null   int64
26  web_traffic                   11055 non-null   int64
27  Page_Rank                     11055 non-null   int64
28  Google_Index                  11055 non-null   int64
29  Links_pointing_to_page        11055 non-null   int64
30  Statistical_report            11055 non-null   int64
31  Result                        11055 non-null   int64
dtypes: int64(32)
memory usage: 2.7 MB
```

data.describe()

In [6]:

Out[6]:

	index	having_IP_having_IP_Address	URL_URL_Length	Shortining_Service	having_At_Symbol	double_slash_redirecting	Prefix_Suffix	having_Sub_Domain	SSLfinal_State	Domain_registeration_length	. .	popUpWidnow	Iframe	age_of_domain	DNSRecord	web_traffic	Page_Rank	Google_Index	Links_pointing_to_page	Statistical_report	Result
count	11055.000000	110 55. 000 000	11055. 00 00 0	11055. 00 00 0	11055. 00 00 0	11055.0 00 00 0	11055. 000 0	11055. 00 00	11055. 000 0	11055. 000 000	. . .	11055. 000 00	11055. 000 00	11055. 000 00	11055. 000 00	11055. 000 00	11055. 000 00	11055. 000 00	11055.000 000	11055. 000 00	11055. 000 00

	index	having_IP having_IP_Address	URL_URL_Length	Shortining_Service	having_At_Symbol	double_slash_redirecting	Prefix_Suffix	having_Sub_Domain	SSLfinal_State	Domain_registration_length	...	popUpWidnow	Iframe	age_of_domain	DNSRecord	web_traffic	Page_Rank	Google_Index	Links_pointing_to_page	Statistical_report	Result
mean	5528.000000	0.313795	-0.633198	0.738761	0.700588	0.741474	-0.734962	0.063953	0.250927	-0.336771	.	0.613388	0.816915	0.061239	0.377114	0.287291	-0.483673	0.721574	0.344007	0.719584	0.113885
std	3191.447947	0.949534	0.766095	0.673998	0.713598	0.671011	0.678139	0.817518	0.911892	0.941629	.	0.789818	0.576784	0.998168	0.926209	0.827733	0.875289	0.692369	0.569944	0.694437	0.993539
min	1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	.	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	-1.000000	-1.000000	-1.000000
25%	2764.500000	-1.000000	1.000000	1.000000	1.000000	1.000000	-1.000000	-1.000000	-1.000000	-1.000000	.	1.000000	1.000000	-1.000000	-1.000000	0.000000	-1.000000	1.000000	0.000000	1.000000	-1.000000
50%	5528.0	1.000000	-1.00	1.00	1.00	1.00	-1.00	0.00	1.00	-1.00	.	1.00	1.00	1.00	1.00	1.00	-1.00	1.00	0.000000	1.00	1.00

	index	having_IPhaving_IP_Address	URLURL_Length	Shortining_Service	having_At_Symbol	double_slash_redirecting	Prefix_Suffix	having_Sub_Domain	SSLfinal_State	Domain_registeration_length	. . .	popUpWidnow	Iframe	age_of_domain	DNSRecord	web_traffic	Page_Rank	Google_Index	Links_pointing_to_page	Statistical_report	Result
	0		00				0		0			0	0	0	0	0	0	0		0	0
	0		00				0		0			0	0	0	0	0	0	0		0	0
	0						0											0			
	0																				
	0																				
75%	829.50000	1.00000	1.0000	1.000	1.000	1.0000	-1.000	1.000	1.0000	1.0000	. . .	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
max	1105.00000	1.0000	1.00000	1.000	1.000	1.0000	1.000	1.000	1.00000	1.000	. . .	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

8 rows × 32 columns

In [7]:

```python
# Handling Null values
# Checking for null values
data.isnull().any()
```

Out[7]:

```
index                          False
having_IPhaving_IP_Address     False
URLURL_Length                  False
Shortining_Service             False
having_At_Symbol               False
double_slash_redirecting       False
Prefix_Suffix                  False
having_Sub_Domain              False
SSLfinal_State                 False
Domain_registeration_length    False
```

```
Favicon                       False
port                          False
HTTPS_token                   False
Request_URL                   False
URL_of_Anchor                 False
Links_in_tags                 False
SFH                           False
Submitting_to_email           False
Abnormal_URL                  False
Redirect                      False
on_mouseover                  False
RightClick                    False
popUpWidnow                   False
Iframe                        False
age_of_domain                 False
DNSRecord                     False
web_traffic                   False
Page_Rank                     False
Google_Index                  False
Links_pointing_to_page        False
Statistical_report            False
Result                        False
dtype: bool
```

```
data.isnull().sum()
```

```
index                         0
having_IPhaving_IP_Address    0
URLURL_Length                 0
Shortining_Service            0
having_At_Symbol              0
double_slash_redirecting      0
Prefix_Suffix                 0
having_Sub_Domain             0
SSLfinal_State                0
Domain_registeration_length   0
Favicon                       0
port                          0
HTTPS_token                   0
Request_URL                   0
URL_of_Anchor                 0
Links_in_tags                 0
SFH                           0
Submitting_to_email           0
Abnormal_URL                  0
Redirect                      0
on_mouseover                  0
RightClick                    0
popUpWidnow                   0
Iframe                        0
age_of_domain                 0
DNSRecord                     0
web_traffic                   0
Page_Rank                     0
Google_Index                  0
Links_pointing_to_page        0
```

```
Statistical_report          0
Result                      0
dtype: int64
```

```python
# Splitting data into independent and dependent variables
x = data.iloc[:,1:31].values
y = data.iloc[:,-1].values
print(x)
print(y)
```

```
[[-1  1  1 ...  1  1 -1]
 [ 1  1  1 ...  1  1  1]
 [ 1  0  1 ...  1  0 -1]
 ...
 [ 1 -1  1 ...  1  0  1]
 [-1 -1  1 ...  1  1  1]
 [-1 -1  1 ... -1  1 -1]]
[-1 -1 -1 ... -1 -1 -1]
```

```python
# Splitting data into train and test
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test =  train_test_split(x, y, test_size = 0.2,
random_state  = 0)
```

```python
# model building
from sklearn.metrics import accuracy_score, classification_report
```

```python
# decision tree
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()
dt.fit(x_train, y_train)
prediction_dt = dt.predict(x_test)
accuracy_dt = accuracy_score(y_test, prediction_dt)*100
scores_dict = {}
```

```python
print('Accuracy score:', accuracy_dt)
scores_dict['DecisionTreeClassifier'] = accuracy_dt
print(classification_report(y_test, prediction_dt))
```

```
Accuracy score: 96.29127091813659
              precision    recall  f1-score   support

          -1       0.97      0.95      0.96      1014
           1       0.96      0.97      0.97      1197

    accuracy                           0.96      2211
   macro avg       0.96      0.96      0.96      2211
weighted avg       0.96      0.96      0.96      2211
```

```python
dt.feature_importances_
```

```
array([0.00749069, 0.00765109, 0.00213221, 0.00269844, 0.00353234,
       0.0190622 , 0.0308616 , 0.62671122, 0.01614264, 0.003762  ,
       0.00067306, 0.00470924, 0.01131087, 0.10812811, 0.03399226,
```

```
        0.01001914, 0.00732687, 0.00225578, 0.00419494, 0.00204161,
        0.00197226, 0.00199053, 0.00234312, 0.01432219, 0.00824365,
        0.02808774, 0.00507387, 0.01050455, 0.01927073, 0.00349503])
```

In [32]:

```python
# Logistic Regression
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
lr.fit(x_train, y_train)
```

Out[32]:

```
LogisticRegression()
```

In [37]:

```python
y_pred1= lr.predict(x_test)
from sklearn.metrics import accuracy_score
log_reg = accuracy_score(y_test, y_pred1)*100
log_reg
```

Out[37]:

```
91.67797376752601
```

In [38]:

```python
scores_dict['LogisticRegression'] = log_reg
```

In [39]:

```python
algo_name = list(scores_dict.keys())
accuracy_list = list(scores_dict.values())
sns.set(rc={'figure.figsize':(12.5,6.5)})
with sns.color_palette('muted'):
    sns.barplot(x=algo_name,y=accuracy_list)
```