



МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ім. ІГОРЯ СІКОРСЬКОГО»  
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ  
КАФЕДРА ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

**Теорія та методи соціальної  
інженерії в кібербезпеці  
Лабораторний практикум №4**

**Виявлення фішингових сайтів  
методами машинного навчання**

Перевірів:  
Кіфорчук К. О.

Виконав:  
студент I курсу  
групи ФБ-41мп  
Сахній Н. Р.

Київ 2025

**Мета роботи:** Ознайомитись із засобами, які можуть використовуватись для виявлення шкочочинності (використання для цілей фішингу) веб-сайта, зокрема на основі різних методів машинного навчання.

## **Завдання до виконання:**

**1.** Відібрати перелік ознак для аналізу із числа характеристик:

*a) HTML та JS*

Current relation

Relation: phishing-weka.filters.unsupervised.attribute.Remove-...  
Instances: 11055

Attributes: 6  
Sum of weights: 11055

Attributes

All

None

Invert

Pattern

No.	Name
1	<input type="checkbox"/> Redirect
2	<input type="checkbox"/> on_mouseover
3	<input type="checkbox"/> RightClick
4	<input type="checkbox"/> popUpWidnow
5	<input type="checkbox"/> Iframe
6	<input checked="" type="checkbox"/> Result

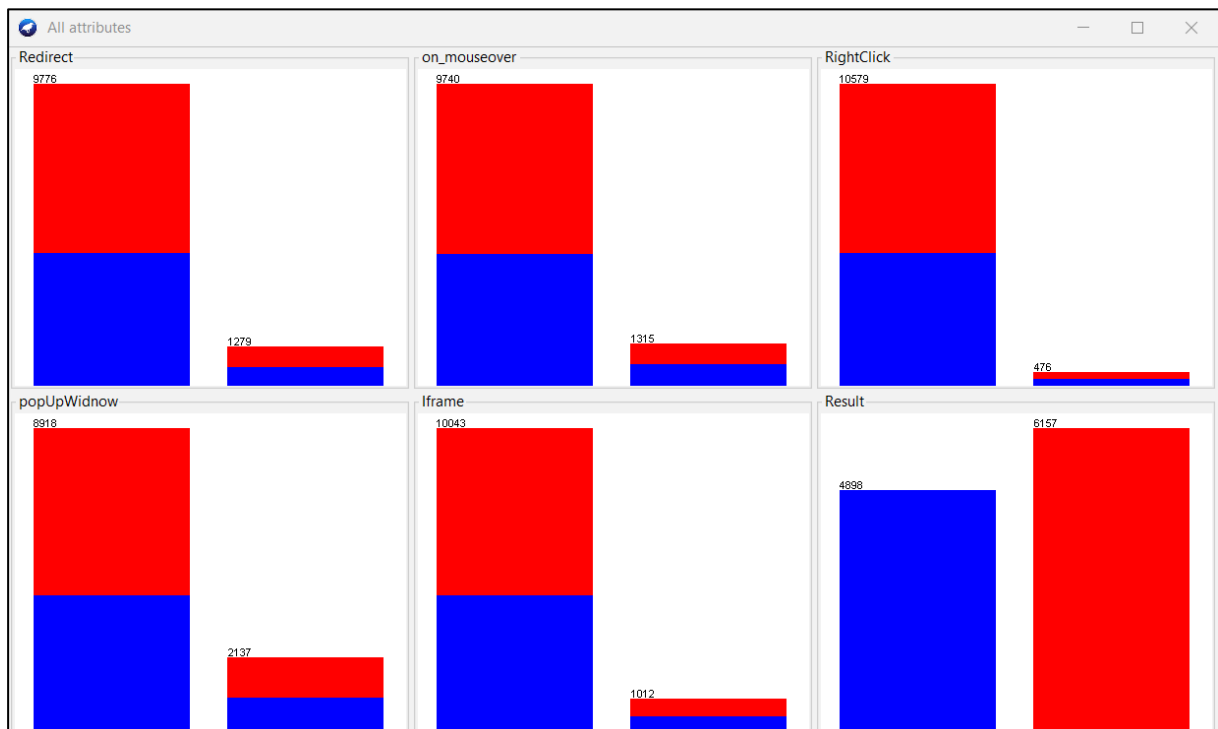
Selected attribute

Name: Result  
Missing: 0 (0%)

Distinct: 2

Type: Nominal  
Unique: 0 (0%)

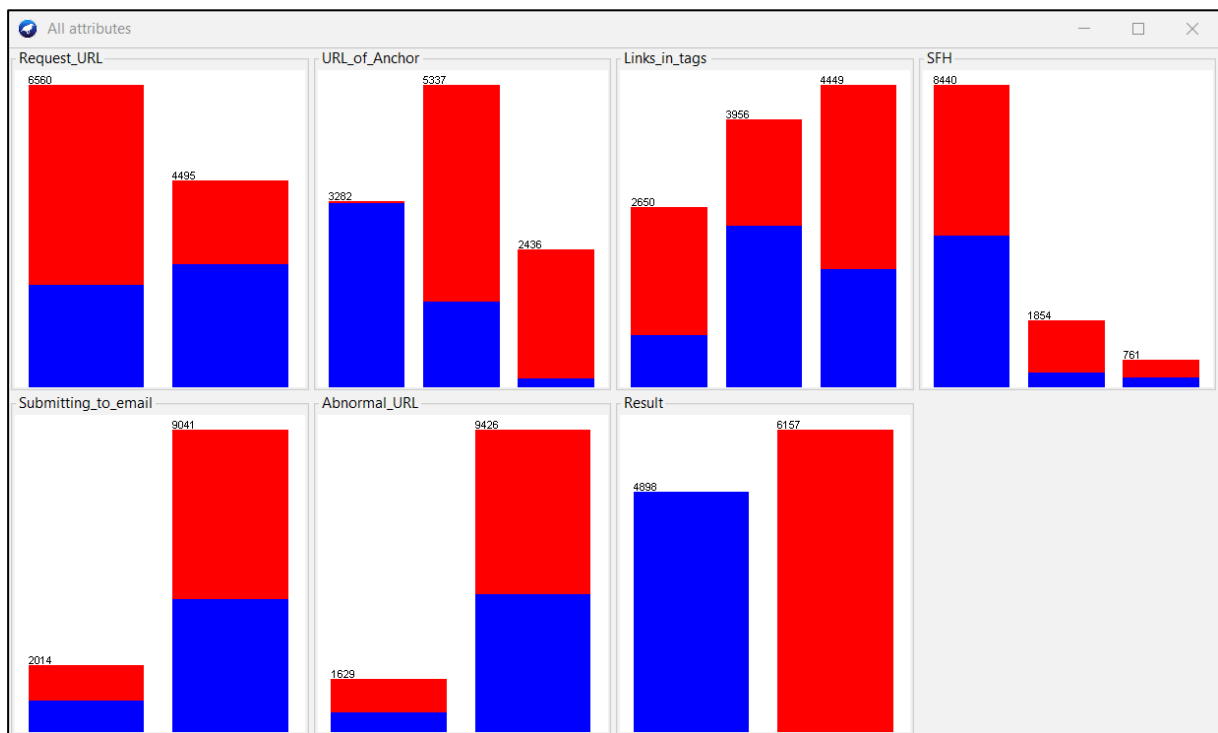
No.	Label	Count	Weight
1	-1	4898	4898
2	1	6157	6157



## б) аномальных характеристик

Current relation		Attributes: 7	
Relation: phishing-weka.filters.unsupervised.attribute.Remove-...		Sum of weights: 11055	
Instances: 11055			
Attributes			
<div>All</div>		<div>None</div>	
<div>Invert</div>		<div>Pattern</div>	
No.	Name		
1	<input type="checkbox"/> Request_URL		
2	<input type="checkbox"/> URL_of_Anchor		
3	<input type="checkbox"/> Links_in_tags		
4	<input type="checkbox"/> SFH		
5	<input type="checkbox"/> Submitting_to_email		
6	<input type="checkbox"/> Abnormal_URL		
7	<input type="checkbox"/> Result		

Selected attribute			
Name: Result		Type: Nominal	
Missing: 0 (0%)		Distinct: 2	
Unique: 0 (0%)			
No.	Label	Count	Weight
1	-1	4898	4898
2	1	6157	6157



## в) характеристик адресного ряда

Current relation

Relation: phishing-weka.filters.unsupervised.attribute.Remove-...  
Instances: 11055

Attributes: 13  
Sum of weights: 11055

Attributes

All

None

Invert

Pattern

No.

Name

1

☐

having\_IP\_Address

2

☐

URL\_Length

3

☐

Shortning\_Service

4

☐

having\_At\_Symbol

5

☐

double\_slash\_redirecting

6

☐

Prefix\_Suffix

7

☐

having\_Sub\_Domain

8

☐

SSLfinal\_State

9

☐

Domain\_registration\_length

10

☐

Favicon

11

☐

port

12

☐

HTTPS\_token

13

☐

Result

Selected attribute

Name: Result  
Missing: 0 (0%)

Type: Nominal  
Unique: 0 (0%)

No.

Label

Count

Weight

1

-1

4898

4898

2

1

6157

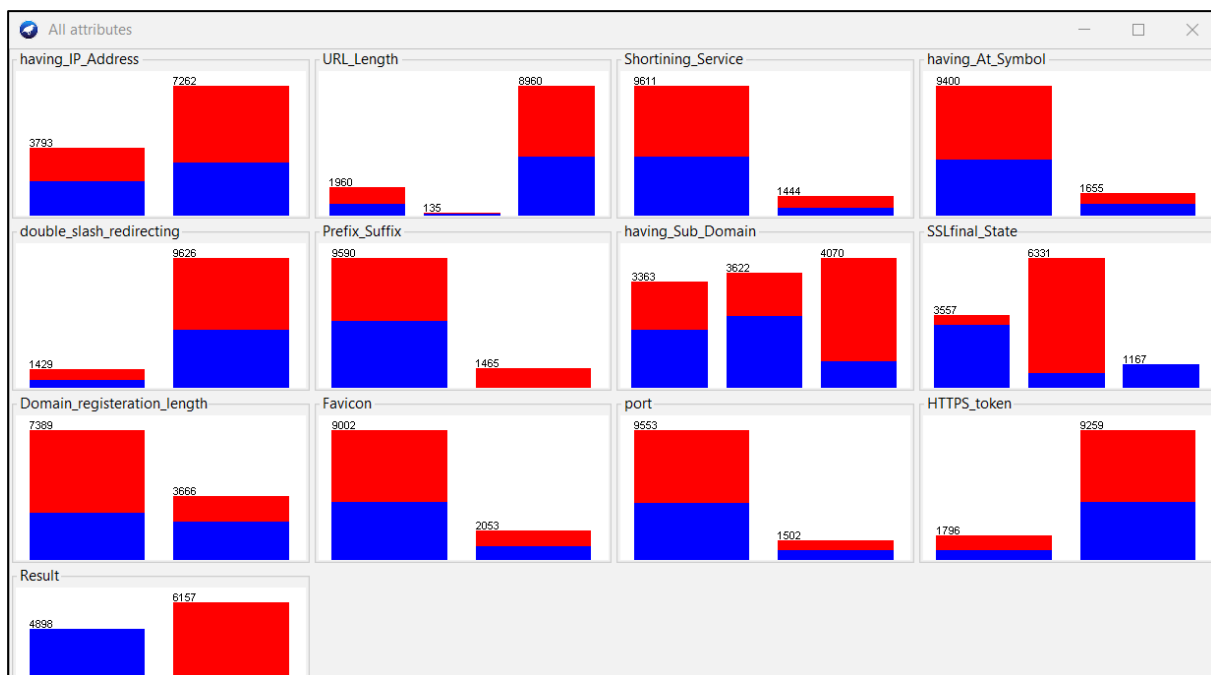
6157

Class: Result (Nom)

Visualize All

4898

6157

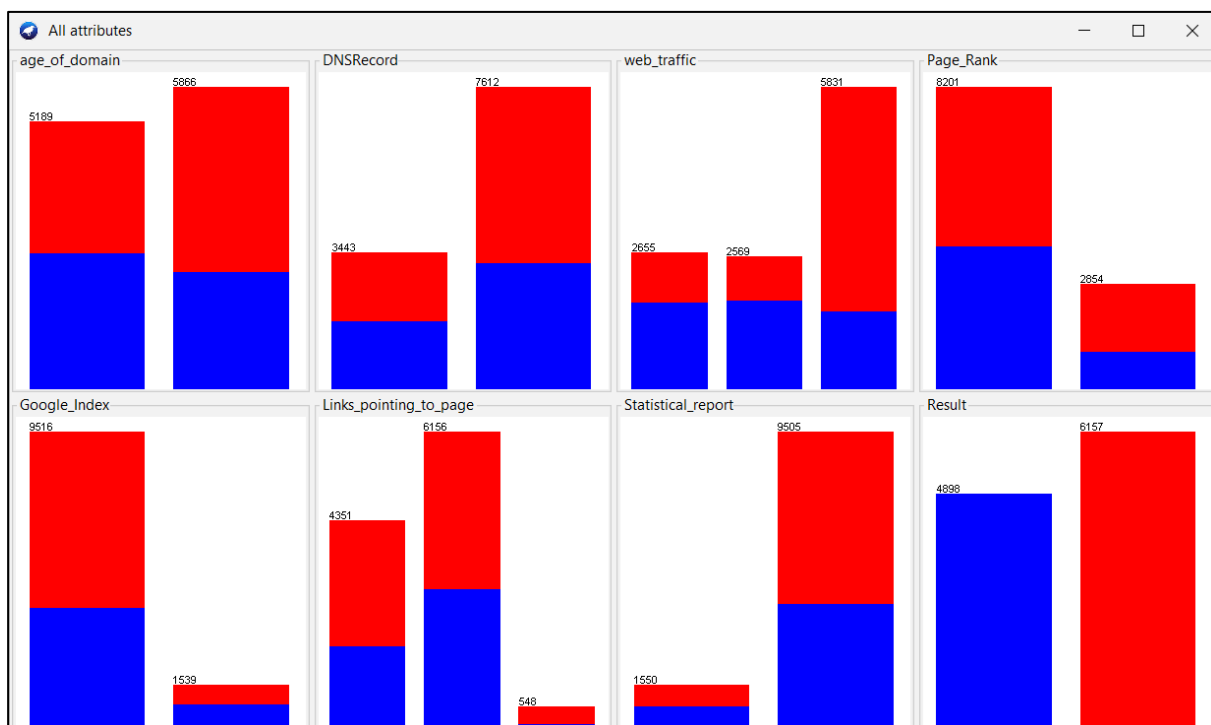


## 2) характеристик домену

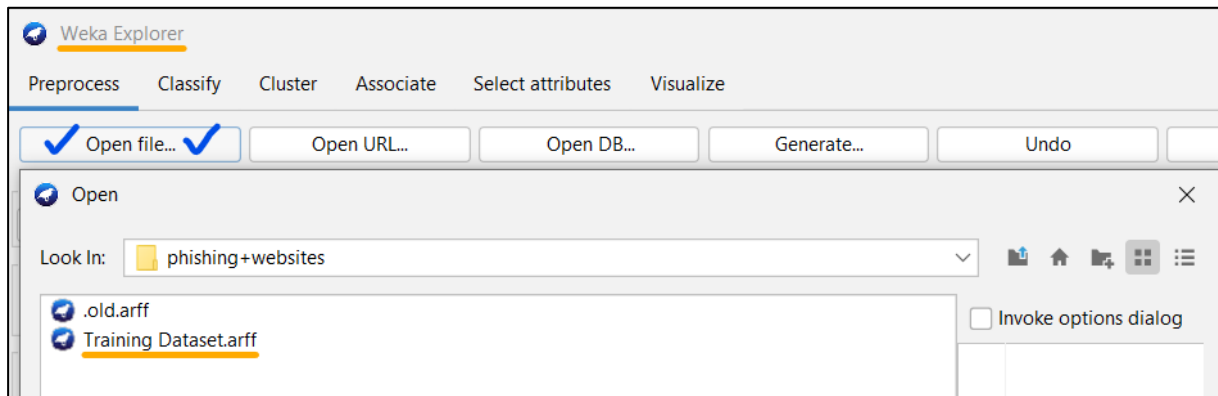
Current relation		Attributes: 8	
Relation: phishing-weka.filters.unsupervised.attribute.Remove-...		Sum of weights: 11055	
Instances: 11055			
Attributes			
<div>All</div>		<div>None</div>	
<div>Invert</div>		<div>Pattern</div>	
No.	Name		
1	<input type="checkbox"/> age_of_domain		
2	<input type="checkbox"/> DNSRecord		
3	<input type="checkbox"/> web_traffic		
4	<input type="checkbox"/> Page_Rank		
5	<input type="checkbox"/> Google_Index		
6	<input type="checkbox"/> Links_pointing_to_page		
7	<input type="checkbox"/> Statistical_report		
8	<input checked="" type="checkbox"/> Result		

Selected attribute			
Name: Result		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
Distinct: 2			
No.	Label	Count	Weight
1	-1	4898	4898
2	1	6157	6157

Class: Result (Nom)		Visualize All
---------------------	--	---------------



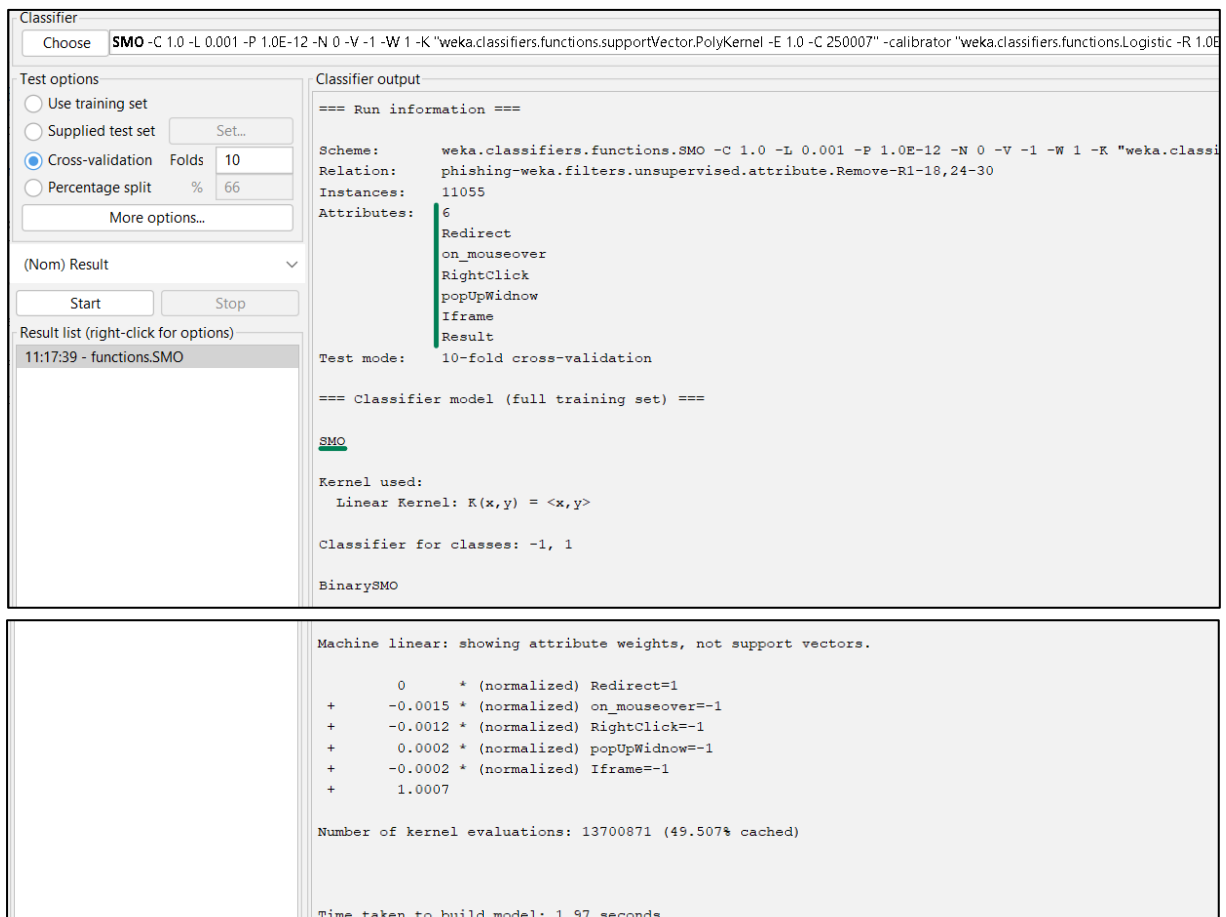
## 2. Конвертувати файл із даними у формат ARFF.



## 3. Використати [класифікатор WEKA](#) в режимі алгоритму, відповідного варіанту \*. В якості вихідних даних використати [датасет з UCI MLR](#).

\*10mod6 = №4. Використати класифікатор Sequential minimal optimization (SMO)

*a) HTML та JS*



## б) аномальных характеристик

Classifier

Choose **SMO** -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E

Test options

☐ Use training set

☐ Supplied test set 

Set...

☒ Cross-validation 

Folds 10

☐ Percentage split 

% 66

More options...

(Nom) Result

Start Stop

Result list (right-click for options)

11:17:39 - functions.SMO

11:27:35 - functions.SMO

Classifier output

=== Run information ===

Scheme: weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classi

Relation: phishing-weka.filters.unsupervised.attribute.Remove-R1-12,19-30

Instances: 11055

Attributes: 7

Request\_URL

URL\_of\_Anchor

Links\_in\_tags

SFH

Submitting\_to\_email

Abnormal\_URL

Result

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

**SMO**

Kernel used:

Linear Kernel:  $K(x,y) = \langle x,y \rangle$

Classifier for classes: -1, 1

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.0009 \* (normalized) Request\_URL=-1

+ -3.3315 \* (normalized) URL\_of\_Anchor=-1

+ 0.667 \* (normalized) URL\_of\_Anchor=0

+ 2.6645 \* (normalized) URL\_of\_Anchor=1

+ 0.6663 \* (normalized) Links\_in\_tags=1

+ -1.3324 \* (normalized) Links\_in\_tags=-1

+ 0.666 \* (normalized) Links\_in\_tags=0

+ -1.3318 \* (normalized) SFH=-1

+ 0.6659 \* (normalized) SFH=1

+ 0.6658 \* (normalized) SFH=0

+ -0.0006 \* (normalized) Submitting\_to\_email=1

+ -0.0006 \* (normalized) Abnormal\_URL=1

+ 1.0002

Number of kernel evaluations: 45129954 (43.483% cached)

Time taken to build model: 6.52 seconds

## в) характеристик адресного ряда

Classifier

Choose **SMO** -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E

Test options

☐ Use training set

☐ Supplied test set 

Set...

☒ Cross-validation 

Folds 10

☐ Percentage split 

% 66

More options...

(Nom) Result

Start Stop

Result list (right-click for options)

11:17:39 - functions.SMO

11:27:35 - functions.SMO

11:35:30 - functions.SMO

Classifier output

=== Run information ===

Scheme: weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classi

Relation: phishing-weka.filters.unsupervised.attribute.Remove-R13-30

Instances: 11055

Attributes: 13

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

Result

Test mode: 10-fold cross-validation

Start

Stop

Result list (right-click for options)

11:17:39 - functions.SMO

11:27:35 - functions.SMO

11:35:30 - functions.SMO

Machine linear: showing attribute weights, not support vectors.

0.0003 \* (normalized) having\_IP\_Address=1

+ 0.0006 \* (normalized) URL\_Length=1

+ -0.0007 \* (normalized) URL\_Length=0

+ 0.0002 \* (normalized) URL\_Length=-1

+ 0.0009 \* (normalized) Shortining\_Service=-1

+ -0.0001 \* (normalized) having\_At\_Symbol=-1

+ 0.0008 \* (normalized) double\_slash\_redirecting=1

+ 2.0003 \* (normalized) Prefix\_Suffix=1

+ -0.0001 \* (normalized) having\_Sub\_Domain=-1

+ -0.0001 \* (normalized) having\_Sub\_Domain=0

+ 0.0002 \* (normalized) having\_Sub\_Domain=1

+ -0.6667 \* (normalized) SSLfinal\_State=-1

+ 1.3335 \* (normalized) SSLfinal\_State=1

+ -0.6668 \* (normalized) SSLfinal\_State=0

+ -0.0002 \* (normalized) Domain\_registration\_length=1

+ -0.0002 \* (normalized) Favicon=-1

+ -0 \* (normalized) port=-1

+ -0.0003 \* (normalized) HTTPS\_token=1

- 0.3342

Number of kernel evaluations: 25184246 (38.702% cached)

Time taken to build model: 4.5 seconds

## 2) характеристик домену

Classifier

Choose SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) Result

Start

Stop

Result list (right-click for options)

11:17:39 - functions.SMO

11:27:35 - functions.SMO

11:35:30 - functions.SMO

11:41:06 - functions.SMO

Classifier output

=== Run information ===

Scheme: weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0

Relation: phishing-weka.filters.unsupervised.attribute.Remove-R1-23

Instances: 11055

Attributes: 8

age\_of\_domain

DNSRecord

web\_traffic

Page\_Rank

Google\_Index

Links\_pointing\_to\_page

Statistical\_report

Result

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

SMO

Kernel used:

Linear Kernel:  $K(x,y) = \langle x,y \rangle$

Classifier for classes: -1, 1

BinarySMO

```
Machine linear: showing attribute weights, not support vectors.

      0.0002 * (normalized) age_of_domain=1
+      0.0001 * (normalized) DNSRecord=1
+     -0.6666 * (normalized) web_traffic=-1
+     -0.6668 * (normalized) web_traffic=0
+      1.3334 * (normalized) web_traffic=1
+      0.0002 * (normalized) Page_Rank=1
+     -0.0001 * (normalized) Google_Index=-1
+      0.0002 * (normalized) Links_pointing_to_page=1
+     -0.0001 * (normalized) Links_pointing_to_page=0
+     -0.0001 * (normalized) Links_pointing_to_page=-1
+      0.0004 * (normalized) Statistical_report=1
-      0.3337

Number of kernel evaluations: 26844120 (40.547% cached)

Time taken to build model: 4.33 seconds
```

*\*) увесь перелік характеристик*

Classifier
Choose **SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"** -calibrator "weka.classifiers.functions.Logistic -R 1.0E

Test options
☐ Use training set
☐ Supplied test set 
☒ Cross-validation Folds 
☐ Percentage split %

(Nom) Result
Result list (right-click for options)

11:17:39 - functions.SMO
11:27:35 - functions.SMO
11:35:30 - functions.SMO
11:41:06 - functions.SMO
11:49:18 - functions.SMO

Classifier output

```

=== Run information ===

Scheme:      weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classi
Relation:    phishing
Instances:    11055
Attributes:   31
             having_IP_Address
             URL_Length
             Shortning_Service
             having_At_Symbol
             double_slash_redirecting
             Prefix_Suffix
             having_Sub_Domain
             SSLfinal_State
             Domain_registration_length
             Favicon
             port
             HTTPS_token
             Request_URL
             URL_of_Anchor

```

```

             Page_Rank
             Google_Index
             Links_pointing_to_page
             Statistical_report
             Result
Test mode:    10-fold cross-validation

```

```

=== Classifier model (full training set) ===

SMO

Kernel used:
  Linear Kernel: K(x,y) = <x,y>

Classifier for classes: -1, 1

BinarySMO

Machine linear: showing attribute weights, not support vectors.

      1.0171 * (normalized) having_IP_Address=1
+      0.104 * (normalized) URL_Length=1
+     -0.1083 * (normalized) URL_Length=0
+      0.0043 * (normalized) URL_Length=-1
+     -0.9642 * (normalized) Shortning_Service=1

```



```

+ 0.127 * (normalized) web_traffic=-1
+ -0.6017 * (normalized) web_traffic=0
+ 0.4747 * (normalized) web_traffic=1
+ 0.1649 * (normalized) Page_Rank=1
+ -0.8231 * (normalized) Google_Index=-1
+ 0.5623 * (normalized) Links_pointing_to_page=1
+ -0.3823 * (normalized) Links_pointing_to_page=0
+ -0.1801 * (normalized) Links_pointing_to_page=-1
+ 0.1828 * (normalized) Statistical_report=1
- 0.8852

Number of kernel evaluations: 63622517 (46.764% cached)

Time taken to build model: 13.66 seconds

```

#### 4. Одержати показники Accuracy, Precision, Recall для характеристик.

a) HTML та JS: (55,4 %)

```

Result list (right-click for options)
11:17:39 - functions.SMO
11:27:35 - functions.SMO
11:35:30 - functions.SMO
11:41:06 - functions.SMO
11:49:18 - functions.SMO

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      6127      55.4229 %
Incorrectly Classified Instances    4928      44.5771 %
Kappa statistic                    0.0067
Mean absolute error                 0.4458
Root mean squared error             0.6677
Relative absolute error             90.3255 %
Root relative squared error        134.4066 %
Total Number of Instances          11055

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,054    0,048    0,473     0,054    0,096      0,014    0,503    0,445    -1
               0,952    0,946    0,559     0,952    0,704      0,014    0,503    0,558     1
Weighted Avg.   0,554    0,548    0,521     0,554    0,435      0,014    0,503    0,508

=== Confusion Matrix ===
      a    b  <-- classified as
263 4635 |   a = -1
293 5864 |   b = 1

```

б) аномальних характеристик – (86,2 %)

```

Result list (right-click for options)
11:17:39 - functions.SMO
11:27:35 - functions.SMO
11:35:30 - functions.SMO
11:41:06 - functions.SMO
11:49:18 - functions.SMO

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      9524      86.1511 %
Incorrectly Classified Instances    1531      13.8489 %
Kappa statistic                    0.7179
Mean absolute error                 0.1385
Root mean squared error             0.3721
Relative absolute error             28.0618 %
Root relative squared error        74.9157 %
Total Number of Instances          11055

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,818    0,104    0,862     0,818    0,840      0,719    0,857    0,786    -1
               0,896    0,182    0,861     0,896    0,878      0,719    0,857    0,829     1
Weighted Avg.   0,862    0,147    0,862     0,862    0,861      0,719    0,857    0,810

=== Confusion Matrix ===
      a    b  <-- classified as
4007 891 |   a = -1
640 5517 |   b = 1

```

в) характеристик адресного рядку – (89,8 %)

```

Result list (right-click for options)

11:17:39 - functions.SMO
11:27:35 - functions.SMO
11:35:30 - functions.SMO
11:41:06 - functions.SMO
11:49:18 - functions.SMO

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      9929      89.8146 %
Incorrectly Classified Instances    1126      10.1854 %
Kappa statistic                    0.7924
Mean absolute error                 0.1019
Root mean squared error             0.3191
Relative absolute error              20.6385 %
Root relative squared error         64.2473 %
Total Number of Instances          11055

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0,857    0,069    0,908      0,857    0,882      0,793    0,894     0,841     -1
                0,931    0,143    0,891      0,931    0,911      0,793    0,894     0,868     1
Weighted Avg.   0,898    0,110    0,899      0,898    0,898      0,793    0,894     0,856

=== Confusion Matrix ===

      a    b  <-- classified as
4197  701 |    a = -1
425  5732 |    b = 1

```

2) характеристик домену – (69,8 %)

```
Result list (right-click for options)
```

11:17:39 - functions.SMO
11:27:35 - functions.SMO
11:35:30 - functions.SMO
11:41:06 - functions.SMO
11:49:18 - functions.SMO

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      7715      69.7874 %
Incorrectly Classified Instances    3340      30.2126 %
Kappa statistic                    0.3919
Mean absolute error                 0.3021
Root mean squared error             0.5497
Relative absolute error             61.219 %
Root relative squared error         110.6518 %
Total Number of Instances          11055

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	FRC Area	Class
	0,692	0,298	0,649	0,692	0,670	0,393	0,697	0,586	-1
	0,702	0,308	0,742	0,702	0,721	0,393	0,697	0,687	1
Weighted Avg.	0,698	0,303	0,701	0,698	0,699	0,393	0,697	0,642	

```
=== Confusion Matrix ===

      a    b  <-- classified as
3391 1507 |  a = -1
1833 4324 |  b = 1
```

\*) увесь перелік характеристик – (93,8 %)

Result list (right-click for options)

```

11:17:39 - functions.SMO
11:27:35 - functions.SMO
11:35:30 - functions.SMO
11:41:06 - functions.SMO
11:49:18 - functions.SMO

```

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      10370           93.8037 %
Incorrectly Classified Instances    685           6.1963 %
Kappa statistic                    0.8742
Mean absolute error                 0.062
Root mean squared error             0.2489
Relative absolute error             12.5554 %
Root relative squared error         50.1107 %
Total Number of Instances          11055

```

=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,920	0,047	0,939	0,920	0,929	0,874	0,936	0,899	-1
	0,953	0,080	0,937	0,953	0,945	0,874	0,936	0,919	1
Weighted Avg.	0,938	0,066	0,938	0,938	0,938	0,874	0,936	0,910	
=== Confusion Matrix ===									
a	b	<-- classified as							
4504	394	a = -1							
291	5866	b = 1							

## 5. Зробити висновок, який із наборів характеристик дає кращі результати.

Посеред всіх типів веб-характеристик найкращим при машинному навчанні із застосуванням алгоритму SMO (Sequential minimal optimization) виявився перелік характеристик адресного рядку (89,8 %). Однак також було проведено навчання із усіма характеристиками різних типів – і воно продемонструвало найвищу точність (93,8 %) за показниками **Accuracy** (відсоток усіх випадків, які модель класифікувала правильно), **Precision** (відсоток із сайтів, класифікованих моделлю як фішингові, що є такими), **Recall** (відсоток із усіх фішингових сайтів у датасеті, які модель змогла правильно визначити).

Це свідчить про те, що комбінування характеристик з різних типів (зокрема, характеристик адресного рядка, структури JS/HTML і т.д.) дозволяє моделі отримати повнішу картину та приймати більш точні рішення при класифікації сайтів.

## Контрольні запитання:

- 1) Які методи машинного навчання можна використовувати для задачі класифікації сайтів на “доброякісні” та “фішингові”?
  - **Naïve Bayes** – кожен параметр класифікованих даних розглядається незалежно від інших параметрів класу. Побудований на теоремі Байеса. Дозволяє передбачити клас використовуючи статистичну ймовірність.

- **Logistic Regression** – статистичний алгоритм бінарної класифікації.
- **J48** – алгоритм будує класифікатор в формі дерева рішень. У кожній точці блок-схеми ставиться питання про значимість тієї чи іншої ознаки, і в залежності від цих ознак екземпляри потрапляють в певний клас.
- **Neural Network** – організовано як розташування взаємопов'язаних нейронів. З'єднання використовуються для відправки сигналів між нейронами.
- **k-Nearest Neighbors** – класифікує об'єкти, порівнюючи їх із найближчими прикладами у навчальному наборі, зазвичай на основі евклідової відстані.
- **Sequential minimal optimization** – алгоритм з групи Support Vector Machine (SVM). Використовує гіперплощину, щоб класифікувати дані по 2 класам. На верхньому рівні SVM виконує ті ж операції, що й алгоритм J48.
- **Ensemble methods** (Random Forest, AdaBoost) – поєднання кількох дерев рішень для досягнення більшої точності і стійкості до шуму.

2) Які підготовчі дії треба виконати з датасетом перед його використанням методами машинного навчання? Яким чином формується подібний датасет перед побудовою моделі?

### **Підготовка до машинного навчання:**

- **Збір даних** – зазвичай з відкритих джерел, таких як PhishTank або OpenPhish. Дані повинні містити як фішингові, так і легітимні URL-адреси сайтів.

- **Формування ознак (фіч)** – витягуються характеристики URL, HTML, JavaScript, домену тощо. Наприклад: довжина URL, використання символів @, форма без дії, кількість редиректів, дата реєстрації домену та т.п.
- **Очищення** – обробка null-значень, видалення дублікатів, фільтрація шуму.
- **Кодування категоріальних ознак** – якщо ознаки мають нечисловий формат (наприклад, тип домену), вони кодуються (наприклад, one-hot encoding).
- **Масштабування/нормалізація** – приведення числових ознак до одного діапазону (наприклад, від 0 до 1) для коректної роботи моделей, особливо kNN.
- **Розбиття датасету** – на тренувальний і тестовий набори (н/д, 80% і 20%).

### **Особливості формування датасету:**

Ознаки мають бути не надто специфічними (наприклад, конкретний текст або бренд), щоб модель не “зазубрила» приклади, а навчилася узагальнювати їх.

3) Які інші методи, окрім методів машинного навчання, можна використовувати для виявлення фішингових сайтів?

- **Аналіз URL-адреси** – виявлення ознак: підозріло довга адреса, наявність IP замість доменного імені, символи @, -, незвичайна структура субдоменів.
- **Чорні списки (Blacklists)** – перевірка URL-адрес у базах відомих фішингових сайтів (наприклад, “Google Safe Browsing”, “PhishTank”). Недоліком таких списків є їх неспроможність захисту від щойно створених фішингових сайтів, оскільки останні спочатку повинні бути виявлені і занесені в списки.

- **Білі списки (Whitelists)** – дозвіл доступу лише до перевірених джерел (наприклад, корпоративних доменів). Ефективно у закритих системах.
- **Евристичні методи** – набір правил (алгоритмів), які вказують на підозрілу активність: відсутність SSL-сертифіката, кількість редиректів, використання `iframe`, тощо. Якщо визначено набір загальних евристичних тестів, то він може бути застосований для виявлення шкідливо створених веб-сайтів.
- **Візуальна схожість** – порівняння зовнішнього вигляду сайту із брендовими зразками (наприклад, фішинговий сайт копіює форму входу Facebook).