Научная статья
Статья в открытом доступе
УДК: 519:004.056.53
doi: 10.30987/2658-4026-2022-4-267-274

Проверка подлинности сайта организации с использованием методов машинного обучения

Александр Николаевич Привалов[1]✉, Вадим Анатольевич Смирнов[2]
[1] Тульский государственный педагогический университет им. Л.Н. Толстого; Тульская область, Тула, Россия
[2] Ивановский государственный университет; Ивановская область, Шуя, Россия
[1] privalov.61@mail.ru; https://orcid.org/0000-0003-3311-0751
[2] v.a.d.i.m@bk.ru; https://orcid.org/0000-0001-6515-9141

Аннотация.
Анализ новостных ресурсов подтверждает наличие финансового ущерба, наносимого пользователям при реализации угроз, источниками которых являются фейковые сайты организации. В качестве решения проблемы фейковых сайтов организации в статье рассматриваются методы проверки подлинности веб-ресурсов. Для этого выделен ряд признаков, характерных для подлинных сайтов организации. Представлена методика формирования выборки подлинных сайтов организации с использованием информационно-аналитических систем для бизнеса. На тренировочной выборке, содержащей сайты из list-org.com и сервиса PhishTank, проведено обучение моделей, основанных на использовании теоремы Байеса, дерева решений и разделяющей гиперплоскости. Приведена гистограмма, отображающая доли сайтов для каждой из выборок, у которых присутствует соответствующий критерий проверки подлинности. При реализации программ, основанных на использовании указанных методов, был использован язык программирования Java и библиотека Weka. Эффективность данных моделей оценена на тестовой выборке, содержащей другие сайты с этих же ресурсов. Разделяющая гиперплоскость позволила обеспечить более высокую общую точность классификации. В то же время при использовании наивного байесовского классификатора было допущено наименьшее количество ошибок, когда фейковые сайты были классифицированы как подлинные.

Ключевые слова: фейковый сайт, подлинность, классификация, дерево решений, наивные байесовские классификаторы, информационная безопасность

Для цитирования: Привалов А. Н., Смирнов В. А. Проверка подлинности сайта организации с использованием методов машинного обучения // Эргодизайн. №4 (18). 2022. С. 267-274. http://dx.doi.org/10.30987/2658-4026-2022-4-267-274.

Original article
Open Access Article

Organising site authentication using machine learning methods

Aleksandr N. Privalov[1]✉, Vadim A. Smirnov[2]
[1] Tula State Pedagogical University named after L.N. Tolstoy; Tula region, Tula, Russia
[2] Ivanovo State University; Ivanovo region, Shuya, Russia
[1] privalov.61@mail.ru; https://orcid.org/0000-0003-3311-0751
[2] v.a.d.i.m@bk.ru; https://orcid.org/0000-0001-6515-9141

Abstract.
The analysis of news resources confirms the presence of financial damage caused to users by threat implications, the sources of which are the organisation's fake websites. As the problem solution of the organisation's fake websites, the article discusses methods for authenticating web resources. For this, a number of features characteristic of the organisation's original sites are identified. The technique of forming a sample of the organization's genuine sites using information and analytical systems for business is presented. On a training sample containing sites from list-org.com and the PhishTank service, models based on applying the Bayes theorem, a decision tree, and a separating hyperplane are carried out. A histogram is provided showing the proportion of sites for each of the samples that have the corresponding authentication criterion. When implementing programmes based on using these methods, the Java programming language and the Weka library are used. The effectiveness of these models is evaluated on a test sample containing other

sites from the same resources. The separating hyperplane allows having a higher overall classification accuracy. At the same time, when using the naive Bayes classifier, the least number of errors are made when the fake sites are classified as genuine.

Keywords: fake site, classification, decision trees, naive bayes, information security

For citation: Privalov A. N., Smirnov V. A. Organising site authentication using machine learning methods // Ergodizayn [Ergodesign], 2022, No. 4 (18). Pp. 267-274. doi:10.30987/2658-4026-2022-4-267-274.

Введение

Одной из самых быстроразвивающихся технологий, используемой в современных компьютерных системах, является технология искусственного интеллекта. Одним из разделов искусственного интеллекта является машинное обучение, характерной чертой которого является попытка прогноза решения некоторой задачи для выбранных объектов на основе известного решения этой же задачи для множества других объектов.

Распространенным вариантом задачи, которая часто решается при помощи методов машинного обучения, является задача классификации объектов по выделенным признакам. Для решения этой задачи специалист должен сформулировать набор признаков, которые будут извлекаться из объектов, и подготовить обучающую выборку. В обучающей выборке должно присутствовать описание признаков объектов и результат их классификации, полученный экспертным путем. Выбранный метод машинного обучения в результате анализа взаимосвязей в этих объектах должен выявить зависимости между значениями признаков и классом объектов. Форма представления данных зависимостей может отличаться и представлять собой: разделяющую гиперплоскость (например, метод опорных векторов (SVM)), дерево решений, параметры для вероятностной модели (наивный байесовский классификатор). Машинное обучение активно применяется в различных областях, требующих принятия решений, в том числе в информационной безопасности. Важность обеспечения информационной безопасности при реализации и развитии платформ и технологий отмечена в правительственной программе «Цифровая экономика» [4].

Одной из актуальных проблем является определение подлинности сайта организации. Согласно исследованию компании Which [2] браузер Google Chrome заблокировал лишь около 25% фишинговых сайтов из представленных в выборке компании. Поддельные сайты становятся инструментом для распространения вредоносного программного обеспечения, к числу которого относится шифровальщик Magniber [6]. Фишинговые кампании, основанные на использовании фейковых сайтов, нередко приводят к краже персональных данных [1], финансовому ущербу для пострадавших [5]. Исходя из актуальности, была сформулирована **цель** исследования – разработать и проверить эффективность критериев проверки подлинности сайта для их использования в методах машинного обучения.

1. Материалы и методы

В исследовании авторов Patil, D., Patil, J. [10] с целью проверки сайта в большей степени используются признаки, извлекаемые из URL-адреса. В их числе: наличие определённых ключевых слов («logon», «signon», «confirm», «account» и др.) в URL-адресе, количество специальных символов («%», «&», «;», «?» и др.). В статье авторов Tubyte, M., Agnė P.-T. [11] анализируются как признаки, извлекаемые из адреса, так и получаемые из внешних источников данные, в том числе: дата регистрации доменного имени (Whois-сервис), посещаемость, наличие в поисковой системе Google и др. В нашем исследовании предполагается выделение признаков подлинности ресурса, основанных на его содержимом.

Для реализации процесса обучения необходимо получить выборку подлинных сайтов организации и фишинговых ресурсов. В качестве общепризнанного источника фишинговых ресурсов в этом исследовании, как и в других, выступает PhishTank.

В качестве выборки подлинных ресурсов возможно использование набора самых популярных сайтов по версии Рейтинг@Mail.ru, LiveInternet и др. Высока вероятность того, что сайты, полученные при помощи выборки адресов с этих ресурсов, будут подлинными. В то же время в ряде случаев эти адреса не будут сайтами организаций. Например, в разделе «Интернет» на 6 месте по популярности на момент написания статьи находится сервис для скачивания видео из социальной сети ТикТок.

Поэтому для получения данных в нашем исследовании была использована одна из информационно-аналитических систем (программно-аппаратный комплекс для сбора, хранения и анализа информации, а затем ее представления в удобном для пользователей виде). Среди таких систем принято выделять интеллектуально-диалоговые системы, системы подготовки принятия решения, статические и

динамические экспертные системы. При принятии бизнес-решений нередко используются такие информационно-аналитические системы, как list-org.com. Данный сервис представляет собой инструмент для анализа деятельности различных компаний. В ряде случае о компании сохраняется ее контактная информация – телефон, факс, e-mail и адрес сайта.

В работе использованы общеизвестные методы машинного обучения. Одним из примененных методов является наивный байесовский классификатор. Его применение основано на теореме Байеса [7]:

$$p(C|F_1,\dots F_K)=\frac{p(C)\,p(F_1,\dots,F_K|C)}{p(F_1,\dots,F_K)},$$

где $p(C)$ – вероятность попадания ресурса в класс подлинных сайтов;

K – количество критериев (т.е. свойства объекта), применяемых при анализе;

F_i – значение i-го критерия из вышеуказанного перечня (при этом $F_i \in \{0,1\}$);

$p(F_1,\dots,F_K)$ – полная вероятность получения значений критериев $F_1,\dots F_K$;

$p(F_1,\dots,F_K|C)$ – вероятность получения значений критериев $F_1,\dots F_K$ для подлинного сайта.

Идея применения наивного байесовского классификатора для принятия решения базируется на предположении о независимости вышеуказанных признаков объектов. Вследствие этого выражение может быть преобразовано следующим образом:

$$p(C|F_1,\dots,F_K)=\frac{p(C)\prod_{i=1}^{K} p(F_i|C)}{p(F_1,\dots,F_K)},$$

где $p(F_i|C)$ – вероятность получения заданного значения F_i для i-го критерия для подлинного веб-ресурса. Вычисление указанных параметров осуществляется в процессе машинного обучения.

Другим способом классификации сайтов является дерево принятия решений. Процесс построения дерева принятия решений заключается в последовательном, рекурсивном выборе правила для разбиения обучающего множества на два подмножества: сайтов, удовлетворяющих указанному правилу, и сайтов, неудовлетворяющих им. Выбор правила (то есть признака для деления на подмножества) может осуществляться различными методами, в том числе основанными на вычислении энтропии, неопределенности Джини, статистической информативности и др.

Метод опорных векторов основан на поиске разделяющей гиперплоскости. Гиперплоскость задается уравнением:

$$w_1 x_1 + w_2 x_2 + \dots + w_K x_K + w_0 = 0,$$

где $w_0 \dots w_K$ – настройки весов, подбираемые в процессе обучения, $x_1 \dots x_K$ – координаты, соответствующие признакам объекта. Класс объектам будет присваиваться исходя из знака результата данного выражения, полученного после подстановки значений свойств объекта.

Для проверки работы методов и сбора данных применялась программа, написанная на языке программирования Java. В работе была использована библиотека для машинного обучения Weka.

2. Результаты

В рамках исследования нами была использована следующая методика формирования выборки подлинных ресурсов, которые являются сайтами организаций:

Шаг 1. Получение исходного кода страницы по URL-адресу https://www.list-org.com/company/**xxxxxx**, где **xxxxxx** – код организации.

Шаг 2. Проверка наличия в коде страницы элемента с классом «warn_red». Этот элемент указывает на факт ликвидации организации. В случае его нахождения возврат к шагу 1 с новым кодом организации.

Шаг 3. Проверка наличия элемента с классом «sites» внутри элемента ... . Если не найден – возврат к шагу 1 с новым кодом организации.

Шаг 4. Сохранение из данного элемента информации о сайте, а из второй ячейки таблицы, расположенной внутри элемента ... – информации о названии организации.

Шаг 5. Возврат к шагу 1 с новым кодом организации. Необходимо учитывать, что между обращениями к сайту должен быть достаточный по продолжительности временной интервал, чтобы избежать повышения нагрузки на сервис list-org.com.

Для формирования репрезентативной выборки подлинных web-ресурсов, в нашем исследовании мы будем брать диапазон кодов организаций (от 1156814 до 1161632).

Несмотря на то, что в результате работы программы будут отобраны только действующие организации, все адреса полученных сайтов необходимо проверить на

работоспособность. В данном исследовании на 4819 страницах с информацией об организациях было найдено 278 адресов сайтов, среди которых только 200 оказались корректными, то есть доменное имя действительно принадлежало организации, информация о которой размещена на сервисе. При этом часть сайтов оказалась недоступна вследствие различных причин (технические работы на сервисе и пр.). В связи с этим для последующего анализа было выбрано 136 работоспособных ресурсов.

С сервиса phishtank были взяты 87 рабочих фишинговых сайтов. При этом под «рабочими сайтами» понимаются те сайты, которые выполняют свою функцию – сбор персональных данных пользователей. Ряд сайтов перестали функционировать вследствие блокировки аккаунта на хостинге, переадресации с доменного имени на другие сайты (с ряда фишинговых сайтов к моменту проведения исследования стала осуществляться переадресация на портал https://2m.ma/ar/). Кроме того, не учитывались сайты, работающие на доменном имени *.weeblysite.com, поскольку доступ к ним из России закрыт.

Далее для каждого из сайтов в обеих выборках были получены значения признаков подлинности сайта. В исследовании были использованы следующие критерии:

1. наличие тега с атрибутом «itemscope» (указывает на наличие микроразметки Schema.org на странице для поисковых систем);
2. наличие тега с атрибутом «property="og:site_name"» (метатег Open Graph с названием сайта – для оформления сниппета у ссылок в социальных сетях);
3. наличие тега с атрибутом «name="google-site-verification"» (метатег для подтверждения прав на сайт в личном кабинете Google Analytics);
4. наличие тега с атрибутом «name="yandex-verification"» (метатег для подтверждения прав на сайт в личном кабинете Яндекс.Вебмастера);
5. наличие тега с атрибутом «name="format-detection"» (метатег, запрещающий мобильному браузеру распознавание номеров телефонов в коде страницы для создания ссылок);
6. наличие тега с атрибутом «type="application/ld+json"» (параметр, указывающий на организацию микроразметки с использованием словаря в JavaScript);
7. наличие тега с атрибутом «itemtype="http://schema.org/Organization"» (указывает на то, что в текущем и вложенных элементах будет присутствовать микроразметка данных о компании);
8. наличие тега с атрибутом «rel="shortcut icon"» (подключение файла с иконкой сайта в браузере);
9. наличие тега с атрибутом «rel="preconnect"» (метатег для ускорения загрузки страницы. Заранее указывает браузеру о необходимости преобразовать имя домена в IP-адрес);
10. наличие тега с атрибутом «rel="alternate"» (метатег для определения языковых вариантов и мобильной версии одной и той же страницы сайта);
11. наличие файла robots.txt (текстовый файл, указывающий параметры индексации сайта);
12. наличие файла sitemap.xml (файл с информацией о страницах сайта, которые необходимо проиндексировать);
13. наличие файла favicon.ico;
14. наличие файла BingSiteAuth.xml (файл подтверждения наличия прав доступа к редактированию сайта для поисковой системы Bing);
15. превышение количества внутренних ссылок на веб-странице над количеством внешних ссылок;
16. наличие на сайте ссылок на социальные сети ВКонтакте и/или Одноклассники;
17. наличие на сайте номера телефона;
18. наличие на сайте адреса электронной почты.

Проверка любого из вышеперечисленных критериев не требует обращения к каким-либо сторонним ресурсам и выполняется при помощи создания GET-запроса на получение файла по URL-адресу или в процессе парсинга исходного кода проверяемой страницы. Результатом проверки веб-ресурса по каждому критерию будет некоторое логическое значение. Более формально, для каждого сайта может быть получено:

$$V = \left(F_1, F_2, \ldots F_K\right),$$

где K – количество критериев ($K = 18$);

F_i – значение i-го критерия из вышеуказанного перечня (при этом $F_i \in \{0, 1\}$).

Доли сайтов для каждой из выборок, у которых присутствует соответствующий критерий проверки подлинности, представлены на рисунке 1.

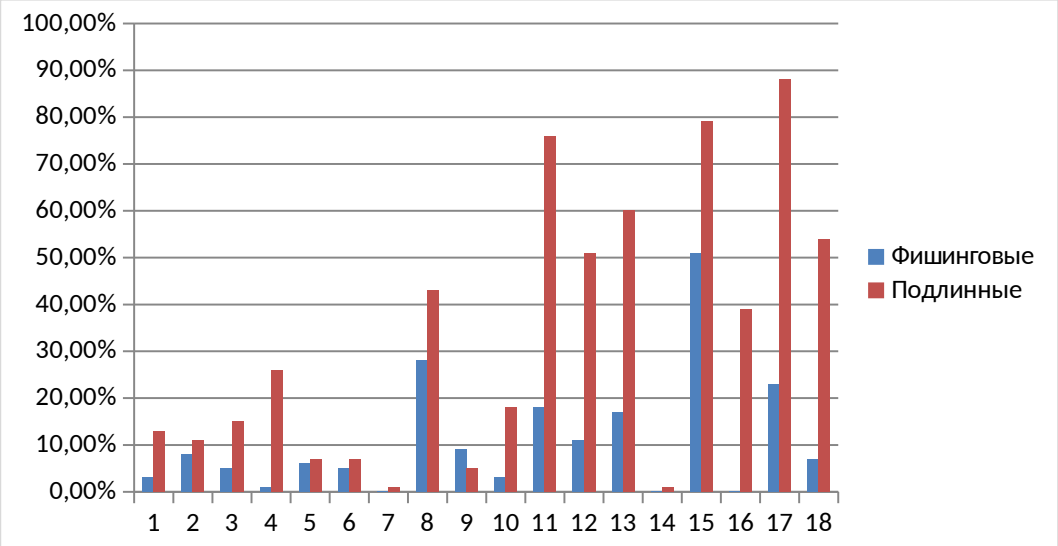

Рис.1. Доли сайтов, обладающих проверяемыми признаками

Fig.1. The proportion of sites with verifiable features

Независимость большинства вышеуказанных признаков позволяет судить о возможности применения наивного байесовского классификатора для принятия решения о подлинности веб-ресурса. Для организации процесса машинного обучения была использована библиотека Weka [12] (класс weka.classifiers.bayes.NaiveBayes) и среда программирования Java.

Тренировка модели выполнялась во фрагменте кода, представленном на рис. 2.

```
22        DataSource source = new DataSource("phishing.arff");
23        Instances train = source.getDataSet();
24        train.setClassIndex(18);
25        NaiveBayes model = new NaiveBayes();
26        model.buildClassifier(train);
```

Рис.2. Код тренировки модели на основе наивного байесовского классификатора на языке программирования Java

Fig. 2. Code for training a model based on a naive Bayesian classifier in the Java programming language

В файле «phishing.arff» представлены результаты анализа фишинговых ресурсов по критериям из начала раздела. Этот файл состоит из двух частей: описания атрибутов объектов и их возможных значений, а также данных, где каждая строка представляет собой описание одного объекта. Содержимое файла представлено на рис. 3.

phishing.arff

```
@RELATION Sites
@ATTRIBUTE p1 {F,T}
@ATTRIBUTE p2 {F,T}
.....
@ATTRIBUTE p18 {F,T}
@ATTRIBUTE R {F,T}
@DATA
F  T  T  T  T  T  F  T  F  F  F  T  T  F  F  F  T  T  T  F
F  F  F  F  F  F  F  F  F  F  F  T  F  T  F  T  T  T  T  F
T  F  F  F  F  F  F  F  T  F  T  T  F  T  F  T  F  T  F  F
F  F  F  T  F  F  F  T  F  T  T  T  T  F  F  T  F  F  F  F
F  F  F  F  F  F  F  T  F  F  F  F  T  F  T  T  T  F  F  F
F  F  F  F  F  F  F  F  F  F  F  F  F  F  T  T  T  T  T  F
F  F  F  F  F  F  F  F  F  F  F  T  F  T  F  F  F  T  F  F
F  F  F  F  F  F  F  T  F  F  F  T  T  T  F  T  T  T  T  F
F  F  F  F  F  F  F  F  F  F  F  T  F  T  F  T  T  T  T  F
T  F  F  F  F  F  F  T  F  F  T  T  T  F  F  T  T  T  T  F
T  F  F  F  F  F  F  T  F  F  T  F  T  F  T  F  T  F  F  F
F  F  F  F  F  F  F  T  F  F  T  T  T  F  F  T  F  T  T  F
F  T  T  T  F  F  F  T  F  T  T  T  F  F  T  F  T  T  T  F
F  F  F  F  F  F  F  F  F  F  T  F  T  F  T  F  T  T  F  F
F  F  F  F  F  F  F  F  F  F  T  T  T  F  T  T  T  T  T  F
F  F  F  F  F  F  F  F  F  F  F  T  F  T  F  T  F  T  F  F
F  F  F  F  F  F  F  T  F  F  F  F  T  F  T  F  T  T  F  F
F  F  F  F  F  F  F  F  F  F  F  T  F  T  F  T  T  F  F  F
F  F  F  T  F  F  F  T  F  F  T  F  T  F  T  F  T  T  F  F
```

Рис. 3. Содержимое файла «phishing.arff»

Fig. 3. Contents of the file "phishing.arff"

Тренировочная выборка составила 105 сайтов. На тестовой выборке из 118 сайтов натренированная модель показала результаты классификации, представленные в таблице 1.

Таблица 1.

Результаты работы наивного байесовского классификатора

Table 1.

Results of the naive Bayesian classifier

		Ответ модели	
		Подлинный	Фишинговый
Мнение экспертов	Подлинный	57 (48,3%)	9 (7,6%)
	Фишинговый	4 (3,4%)	48 (40,7%)

Таким образом, точность прогноза составила 89 %.

Кроме этого была предпринята попытка использования в программном средстве метода построения дерева принятия решений на основе алгоритма C4.5 [9]. Автором данного алгоритма является Джон Квинлан. В Weka реализация этого алгоритма присутствует в виде класса weka.classifiers.trees.J48 (код представлен на рис. 4).

```
51        DataSource src = new DataSource("phishing.arff");
52        Instances dt = src.getDataSet();
53        dt.setClassIndex(18);
54        J48 mytree = new J48();
55        mytree.buildClassifier(dt);
```

Рис. 4. Код тренировки модели на основе C4.5 на языке программирования Java

Fig. 4. C4.5-based model training code in Java programming language

Полученное в результате построения дерево решений изображено на рис. 5. В листьях этого дерева находится значение «T» в случае, если сайт является фейковым, «F» – если подлинным. Надписи у ребер дерева отвечают за выполнение (T) или невыполнение (F) критерия. Номер проверяемого критерия записан в вершине, из которой исходят эти ребра.

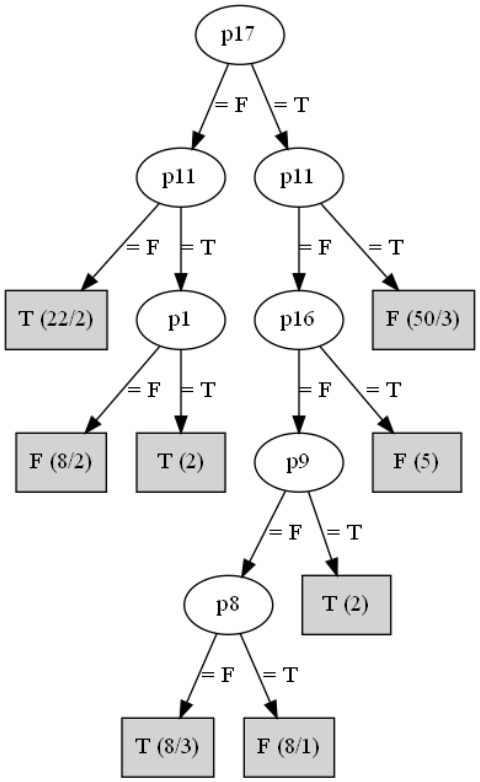

Рис. 5. Дерево принятия решений

Fig. 5. Decision tree

На той же тестовой выборке модель на основе дерева принятия решений показала результаты классификации, представленные в таблице 2.

Точность прогноза составила 82,2 %, что является худшим результатом в сравнении с другими методами. Визуализация модели в виде дерева принятия решений подтверждает склонность модели к переобучению [8]. В

частности, наличие тега с атрибутом «itemscope» в данной модели было воспринято как признак того, что сайт является фейковым (в результате анализа 10 конкретных сайтов, попавших в узел «p1»), хотя более логичным вариантом было бы считать его признаком подлинности. В проанализированных нами сходных исследованиях зарубежных авторов, где используется дерево принятия решений для анализа вредоносности web-ресурса, визуальное отображение графа приведено не было. В связи с этим на данный момент нельзя дать однозначную оценку целесообразности его использования.

Для применения метода опорных векторов [3] был использован класс weka.classifiers.functions.SMO. Код, использованный в программе на Java, представлен на рис. 5.

На той же тестовой выборке построенная модель показала результаты классификации, представленные в таблице 3.

Таким образом, точность прогноза составила 90,7 %.

Таблица 2.

Результаты работы классификатора на основе дерева принятия решений

Table 2.

Results of the classifier based on the decision tree

		Ответ модели	
		Подлинный	Фишинговый
Мнение экспертов	Подлинный	55 (46,6%)	11 (9,3%)
	Фишинговый	10 (8,5%)	42 (35,6%)

```
137         DataSource source = new DataSource("phishing.arff");
138         Instances train = source.getDataSet();
139         train.setClassIndex(18);
140         SMO model = new SMO();
141         model.buildClassifier(train);
```

Рис. 6. Код тренировки модели с использованием метода опорных векторов на языке программирования Java
Fig. 6. Model training code using the support vector machine in the Java programming language

Таблица 3.

Результаты работы классификатора на основе разделяющей гиперплоскости

Table 3.

Results of the classifier based on the separating hyperplane

		Ответ модели	
		Подлинный	Фишинговый
Мнение экспертов	Подлинный	61 (51,7%)	5 (4,2%)
	Фишинговый	6 (5,1%)	46 (39,0%)

Обсуждение/Заключение

При анализе результатов тренировки моделей машинного обучения стоит учитывать, что ошибка принятия подлинного сайта за фишинговый будет менее дорогостоящей, чем классификация фишингового сайта как подлинного. Поэтому, несмотря на более высокую точность классификации, обеспечиваемую гиперплоскостью, в программного средстве проверки подлинности более эффективным будет использование наивного байесовского классификатора.

Исследование может быть расширено путем накопления большей по объему выборки фейковых сайтов и сайтов организаций. Могут быть дополнены и критерии, применяемые при оценке подлинности web-ресурса (за счет введения критериев на основе данных в Whois, сведений из поисковых систем и других источников).

СПИСОК ИСТОЧНИКОВ

1. **Бесплатного пива не бывает: Heineken предупредила о фишинговой кампании в WhatsApp.** 2022. URL: https://www.securitylab.ru/news/532315.php (дата обращения: 17.06.2022).

2. **Браузер Chrome оказался безоружен перед опасными сайтами.** 2022. URL: https://www.gazeta.ru/tech/news/2022/05/30/17835500.shtml (дата обращения: 14.06.2022).

3. **Гончаров Ю.В., Мучник И.Б., Шварцер Л.В.** Алгоритм выбора признаков в задаче обучения классификации методом опорных векторов // Журнал вычислительной математики и математической физики. 2008. Т. 48. № 7. С. 1318-1336.

REFERENCES

1. **There is no Free Beer: Heineken Warned about a Phishing Campaign on WhatsApp [Internet].** 2022 [cited 2022 Jun 17]. Available from: https://www.securitylab.ru/news/532315.php.

2. **Chrome Browser Turned out to Be Unarmed in front of Dangerous Sites [Internet].** 2022 [cited 2022 Jun 14]. Available from: https://www.gazeta.ru/tech/news/2022/05/30/17835500.shtml.

3. **Goncharov Yu.V., Muchnik I.B., Shvartser L.V.** Feature Selection Algorithm in Classification Learning Using Support Vector Machines. Comput. Math. Math. Phys. 2008;48(7):1318-1336.

4. **Ерохин Д.В., Кротенко Т.Н.** Цифровизация экономики в постиндустриальном обществе с позиций институциональных и технологических изменений // Эргодизайн. 2020. № 4(10). С. 177-185. DOI 10.30987/2658-4026-2020-4-177-185.

5. **Как потерять $1,5 миллиона в один клик: мошенник украл 29 NFT-токенов Moonbird с помощью фишинговой ссылки.** Режим доступа: свободный. 2022. URL: https://www.securitylab.ru/news/532019.php (дата обращения: 17.06.2022).

6. **Новая версия вымогательского ПО Magniber угрожает миллионам пользователей Windows 11.** 2022. URL: https://www.securitylab.ru/news/531987.php (дата обращения: 15.06.2022).

7. **Асминг В.Э., Кременецкая Е.О., Виноградов Ю.А. и др.** О применении наивных байесовских классификаторов в сейсмологии // Сейсмические приборы. 2015. Т. 51. № 4. С. 29-40.

8. **Пальмов С.В., Мифтахова А.А.** Реализация деревьев решений в различных аналитических системах // Перспективы науки. 2015. № 1(64). С. 93-98.

9. **Caluza L.J.** Development of J48 Algorithm-Based Application in Predicting Teacher's Techno-Pedagogical Competence // Mindanao Journal of Science and Technology. 2020;18(2):293-310.

10. **Patil D., Patil J.** Malicious URLs Detection Using Decision Tree Classifiers and Majority Voting Technique. Cybernetics and Information Technologies. 2018;18(1):11-29. DOI 10.2478/cait-2018-0002.

11. **Tubyte M., Agnė P.-T.** Research on Phishing Email Detection Based on URL Parameters Using Machine Learning Algorithms // Proceedings of the 26th International Conference on Information Society and University Studies (IVUS 2021). 2021;2915:18-26.

12. **Weka 3 - Data Mining with Open Source Machine Learning Software in Java.** URL: https://www.cs.waikato.ac.nz/ml/weka/ (дата обращения: 06.06.2022).

4. **Erokhin D.V., Krotenko T.N.** Digitalization of the Economy in a Post-Industrial Society from the Perspective of Institutional and Technological Changes. Ergodesign. 2020;4(10):177-185. DOI 10.30987/2658-4026-2020-4-177-185.

5. A **Malicious Link Netted a Scammer $1.5 Million Worth of Moonbird NFTs [Internet].** 2022 [cited 2022 Jun 17]. Available from: https://www.securitylab.ru/news/532019.php.

6. **New Version of Magniber Ransomware Threatens Millions of Windows 11 Users [Internet].** 2022 [cited 2022 Jun 15]. Available from: https://www.securitylab.ru/news/531987.php.

7. **Asming V.E., Kremenetskaya E.O., Vinogradov Yu.A. [et al.]** On Usage of Naive Bayesian Classifiers in Seismology. Seismic Instruments. 2015;51(4):29-40.

8. **Palmov S.V., Miftakhova A.A.** Implementation of Decision Trees in Various Analytical Systems. Science Prospects. 2015;1(64):93-98.

9. **Caluza L.J.** Development of J48 Algorithm-Based Application in Predicting Teacher's Techno-Pedagogical Competence. Mindanao Journal of Science and Technology. 2020;18(2):293-310.

10. **Patil D., Patil J.** Malicious URLs Detection Using Decision Tree Classifiers and Majority Voting Technique. Cybernetics and Information Technologies. 2018;18(1):11-29. DOI 10.2478/cait-2018-0002.

11. **Tubyte M, Agnė P.-T.** Research on Phishing Email Detection Based on URL Parameters Using Machine Learning Algorithms. In: Proceedings of the 26th International Conference on Information Society and University Studies (IVUS 2021). 2021;2915:18-26.

12. **Weka 3 – Data Mining with Open Source Machine Learning Software in Java [Internet]** [cited 2022 Jun 6]. Available from: https://www.cs.waikato.ac.nz/ml/weka/.

Информация об авторах:

Привалов Александр Николаевич - профессор, доктор технических наук, профессор кафедры информатики и информационных технологий, Тульский государственный педагогический университет им. Л.Н. Толстого, международный идентификационный номер автора: SPIN-код: 1016-9572, AuthorID: 522448

Смирнов Вадим Анатольевич - аспирант кафедры математики, информатики и методики обучения, Ивановский государственный университет, международный идентификационный номер автора: SPIN-код: 9378-8141, AuthorID: 961059

Information about the authors:

Privalov Aleksandr Nikolaevich - Professor, Doctor of Technical Sciences, Professor of the department of Informatics and Information Technology, Tula State Pedagogical University named after L.N. Tolstoy, international identification number of the author: Author-ID-RSCI 1016-9572

Smirnov Vadim Anatolyevich - post-graduate student of the Department of Mathematics, Informatics and Teaching Methods, Ivanovo State University; the author's international identification number: SPIN-code: 9378-8141, AuthorID: 961059

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.
Contribution of the authors: the authors contributed equally to this article.
Авторы заявляют об отсутствии конфликта интересов.
The authors declare no conflicts of interests.
Статья поступила в редакцию 02.08.2022; одобрена после рецензирования 09.08.2022; принята к публикации 10.08.2022. Рецензент – Казаков Ю.М., кандидат технических наук., доцент, доцент Брянского государственного технического университета, член редсовета журнала «Эргодизайн».
The paper was submitted for publication on the 2nd of August, 2022; approved after the peer review on the 9th of August, 2022; accepted for publication on the 10th of August, 2022. Reviewer – Kazakov Yu.M., Candidate of Technical Sciences, Associate Professor, Associate Professor of Bryansk State Technical University, member of the editorial board of the journal "Ergodesign".