Вопросы безопасности
Правильная ссылка на статью:
Стрижков В.А. — Применение методов машинного обучения для противодействия инсайдерской угрозе информационной безопасности // Вопросы безопасности. – 2023. – № 4. – С. 152 - 165. DOI: 10.25136/2409-7543.2023.4.68856 EDN: JZMHXQ URL: https://nbpublish.com/library_read_article.php?id=68856

Применение методов машинного обучения для противодействия инсайдерской угрозе информационной безопасности

Стрижков Владислав Александрович

аспирант, Департамент информационной безопасности, Финансовый университет при Правительстве Российской Федерации

125167, Россия, г. Москва, пр-кт Ленинградский, 49/2

218668@edu.fa.ru

Статья из рубрики "Научно-техническое обеспечение национальной безопасности"

DOI:

10.25136/2409-7543.2023.4.68856

EDN:

JZMHXQ

Дата направления статьи в редакцию:

31-10-2023

Дата публикации:

31-12-2023

Аннотация: Предметом исследования является проблема внутренней угрозы информационной безопасности в организациях в лице злонамеренных инсайдеров, а также халатных работников. Объектом исследования являются алгоритмы машинного обучения в части их применимости для обнаружения аномального поведения работников. Автор углубляется в проблематику инсайдерской угрозы, а также рассматривает различные подходы выявления вредоносных действий пользователей, адаптируя эти концепции под наиболее подходящие по функционалу алгоритмы машинного обучения, реализуемые далее в рамках эксперимента. Акцент делается на недостаточности существующих общепринятых мер и политик безопасности и необходимости их усовершенствования за счет новых технологичных решений.

Основным результатом проведённого исследования является опытная демонстрация того, как контролируемое машинное обучение и интеллектуальный анализ данных могут эффективно использоваться для выявления внутренних угроз. При проведении эксперимента используется реалистичный набор входных данных, составленный на основе реальных случаев инсайдерской активности, что позволяет оценить работу алгоритмов машинного обучения в условиях близких к боевым. При сравнении полученных результатов определяется самый эффективный алгоритм, предпочтительный для будущих исследований с более крупным набором данных. Особым вкладом автора выступает свежий взгляд на понимание инсайдерской угрозы и экспериментально обоснованная аргументация в пользу нового подхода противодействия этой угрозе, сочетающего в себе комплекс разноплановых мероприятий. Так, в работе задействуются как математические методы, на которых строится логика машинно-обучаемых алгоритмов: классификация, регрессия, адаптивное повышение и др., так и лингвистические методы, используемые для предварительной обработки входного набора данных, такие как стемминг, векторизация и токенизация.

Ключевые слова:

инсайдерская угроза, внутренний нарушитель, машинное обучение, алгоритмы контролируемого обучения, адаптивное повышение, аномальное поведение, алгоритмы классификации, логистическая регрессия, векторизация, безопасность информации

1. Введение

В современном мире инсайдеры могут оказаться серьезной угрозой для организации, в которой они работают, поэтому предотвращение злонамеренных действий инсайдеров в организационной системе является важной задачей для кибербезопасности. Организационные меры безопасности хорошо известны опытному инсайдеру, и он может легко найти лазейки, а существующие в организациях политики безопасности ориентированы в большей степени на внешние угрозы, порой оставляя без должного внимания возможность появления внутренних угроз. Если необходимые меры безопасности в компании не реализованы, инсайдер получает пространство для кражи важных данных без специфических препятствий, чем наносит непоправимый ущерб организации, как прямой материальный, так и косвенный имиджевый [1]. В связи с этим в текущих реалиях, когда периметр защиты от хакерских и прочих внешних угроз уже выстроен на достаточно высоком и технологичном уровне, стоит на некоторое время сместить акцент с внешних угроз на внутренние, придать данной проблеме значимость и найти новые пути решения. Однако, сложности часто связаны с тем, что реальных статистических данных об инсайдерской угрозе и соответствующем ущербе оказывается мало [2].

Инсайдерскую угрозу можно условно разделить на три основные категории, с которыми регулярно приходится сталкиваться компаниям: злонамеренный инсайдер, «небрежный» инсайдер и скомпрометированный инсайдер. Злонамеренный инсайдер — это тип инсайдера, который намеренно хочет похитить защищаемые данные, раскрыть информацию или с помощью любых других средств нанести вред организации. Инсайдерская угроза «по небрежности» или незлонамеренный инсайдер возникает, когда работники не знают правил безопасности или не соблюдают процедуры безопасности, подвергая компанию риску заражения вредоносным программным

обеспечением и раскрытия данных, становясь таким образом непредумышленными внутренними нарушителями. Наконец, скомпрометированная инсайдерская угроза — это злонамеренный инсайдер, чьи учетные данные были взломаны хакером с помощью социальной инженерии, сбора учетных данных, фишинговых сообщений по электронной почте или методов, использующих уязвимость для кражи данных или с помощью незаконных финансовых транзакций.

Наряду с кибернетическим и техническим сценарием, одной из наиболее важных задач является идентификация и внедрение поведенческих (социотехнических) индикаторов риска инсайдерской угрозы. Когда мы смотрим на данные, доступные для проверки поведения инсайдерской угрозы, зачастую мы игнорируем человеческую сторону вопроса [3]. Эта проблема исследуется специалистами на протяжении многих лет, и в процессе поиска решения уже есть заметные успехи. Один из них – анализ текстов с помощью алгоритмов машинного обучения. Кроме того, имеются данные о том, что алгоритмы обучения с учителем обеспечивают лучшее запоминание по сравнению с полу-контролируемыми и неконтролируемыми методами, когда они получают больше деталей [4].

В данной статье отдельное внимание уделяется категоризации электронной почты из специального набора данных (Далее – НД), который содержит реалистичные примеры инсайдерских угроз, основанных на геймифицированном соревновании, полученных с использованием различных технологий машинного обучения. Во-первых, для очистки данных используются методы предварительной обработки, такие как стемминг, удаление стоп-слов и токенизация. После этого к набору данных для получения информации применяются алгоритмы обучения с учителем Adaboost, а также алгоритм Байеса, логистическая и линейная регрессия, метод опорных векторов. Результаты, полученные с помощью этих алгоритмов, сравниваются анализируются.

Исследование разделено на несколько разделов. Существующие подходы, связанные с этим исследованием, представлены в разделе 2. Информация о наборе данных и методология обсуждается в разделе 3. Результаты исследования представлены в разделе 4. Предстоящая работа и заключительные замечания обсуждаются в пятом разделе.

2. Существующие подходы

Научное сообщество сталкивается с краеугольной проблемой обнаружения инсайдерской угрозы. На ту же проблему ссылается государственный сектор и службы по кибербезопасности, призывая искать эффективные стратегии решения. Ниже представлены имеющиеся работы учёных, посвященные затронутому вопросу. Разумеется, их существует множество, в данном случае отобраны те, в которых задействованы технологии машинного обучения, а потому они заслуживают отдельного внимания, так как могут быть в дальнейшем использованы для противопоставления результатам текущего исследования.

Диоп Абдулай и др. [5] предложили модель обнаружения аномалий поведения, которая является этапом на пути к сборке пользовательской системы проверки поведения. Этот метод объединяет метод классификации машинного обучения и методы на основе графов, основанные на процедурах линейной алгебры и параллельных вычислений. Точность их модели составляет 99% на репрезентативном наборе данных контроля доступа.

Дюк Ли и др.[6] предложили интеллектуальную систему, ориентированную на пользователя, основанную на машинном обучении. Для тщательного мониторинга инсайдера вредоносное поведение пользователей отслеживалось с помощью детального анализа машинного обучения в реальных условиях. Чтобы обеспечить фактическую аппроксимацию производительности системы, предоставляется всесторонний анализ популярных сценариев внутренних угроз с несколькими показателями производительности. Результаты исследования свидетельствуют о том, что система мониторинга и обнаружения на основе искусственного интеллекта (ИИ) может обнаруживать новые внутренние угрозы на основе немаркированных данных с высокой точностью, поскольку она может быть очень эффективно обучена. В исследовании говорится, что примерно 85% таких инсайдерских угроз успешно идентифицированы с показателем ложных срабатываний всего 0,78%.

Мальвика Сингх и др.[7] осуществляли обнаружение инсайдерской угрозы путем работы над другим подходом – профилированием поведения пользователей для наблюдения и изучения последовательности действий поведения пользователей. Исследователи представили гибридную модель машинного обучения, состоящую из «сверхточных нейронных сетей (CNN)» для выявления стабилизирующего выброса в поведенческих паттернах, добившись точности обнаружения 0,9042 и 0,9047 на обучающих и тестовых данных соответственно.

Вслед за этим Цзю мин Лу и др.[8] предложили систему, основанную на глубокой нейронной сети с длинной краткосрочной памятью (LSTM), названную «Ловец инсайдеров», для обучения системных записей файлов, которые действуют как естественная структурированная последовательность. Чтобы отличить нормальное поведение от вредоносного, их система использует шаблоны, которые определяют поведение пользователя. Эксперименты и результаты показали, что предложенный метод продемонстрировал лучшую производительность, чем повсеместно распространённая система обнаружения аномальных ситуаций на основе журналов. Производительность этой системы достаточна для онлайн-сценариев в реальном времени.

Наконец, можно отметить подход Гавай и др.[9], в котором используется «Лес изоляции» - неконтролируемый алгоритм обучения для обнаружения аномалий и выявления инсайдерской угрозы по журналам, взятым из сети.

3. Методология исследования

В этом разделе мы реализуем некоторые методы, которые были адаптированы для анализа данных. Это включает в себя определение набора данных, такие методологии, как методы машинного обучения и обнаружение аномальных электронных писем. На рисунке 1 показана краткая характеристика нашей системы, которая состоит из четырех блоков: блок сбора и предварительной обработки данных, блок преобразования данных, блок контролируемого обучения и блок классификации соответственно. Каждый блок подробно описан в следующих подразделах.

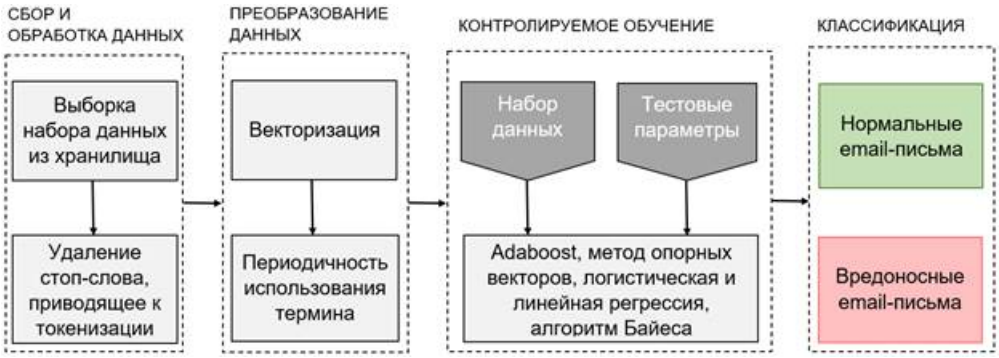

Рис. 1. Векторное представление текстовых данных

3.1. Набор данных

Набор данных (НД) был собран на основе фактического взаимодействия пользователя с хост-машиной, который содержит как легитимные пользовательские данные, так и вредоносные инсайдерские экземпляры (маскарадеры и предатели) [10]. НД содержит данные, полученные из шести источников данных (нажатия клавиш, мышь, монитор хоста, сетевой трафик, электронная почта и вход в систему), а также дополнительные результаты опросника психологической личности. НД содержит поведение 24 пользователей, которые были собраны в течение 5 дней. Он содержит двенадцать экземпляров маскарада, каждый продолжительностью 90 минут, и пять потенциальных экземпляров предателя, каждый из которых длится 120 минут. Было обнаружено, что электронные письма являются важным атрибутом для обнаружения внутренних угроз. В порядке отправки писем пользователями активность всех пользователей содержится в одном файле. Он включает в себя такую информацию, как временная метка, заголовок, отправитель, получатель, функции, извлеченные из тела электронного письма.

3.2. Машинное обучение

Этот метод позволяет выявить закономерности опасного поведения внутренних нарушителей. В статье рассматриваются следующие алгоритмы обучения:

1) Контролируемое обучение – может быть плодотворно, когда известен подтверждающий ответ. Все электронные письма из набора данных классифицируются как «Обычная электронная почта» или «Аномальная электронная почта». Задачи обучения с учителем делятся на две категории: «Классификация» и «Регрессия».

2) Классификация – результаты прогнозируются в дискретном выходе. Входные переменные разделены на дискретные категории. Алгоритмы классификации:

· Алгоритм K-ближайших соседей исследует весь набор данных для k числа наиболее однородных или смежных случаев, которые указывают на ту же модель, что и строка с потерянными данными. Элемент отбрасывается большинством голосов его соседей, элемент присваивается наиболее распространенному классу среди ближайших соседей k [11].

· Алгоритм Байеса – это простой алгоритм обучения, который использует правило Байеса одновременно с жестким предположением о том, что для данного класса атрибуты самодостаточны. Несмотря на то, что этой презумпцией свободы часто злоупотребляют в повседневной жизни, тем не менее, Байес по-прежнему обеспечивает точность для конкурентной классификации [12, 13].

· Метод опорных векторов представляет собой линейную модель для классификации. Алгоритм создает линию или гиперплоскость, которая разделяет данные на классы. [14].

3) Регрессия – результаты прогнозируются в непрерывном выходе. Входные переменные сопоставляются с непрерывной функцией. Алгоритмы регрессии:

· Линейная регрессия – заключается в нахождении наиболее эффективной прямой линии поперек линий. Линия, которая подходит лучше всего, называется линией регрессии. Алгоритм линейной регрессии изучает линейную функцию $f(x, w)$, которая является отображением $f : \varphi x \rightarrow y$. Она представляет собой линейную комбинацию фиксированного набора линейных или нелинейных функций входной переменной, обозначаемых как $\varphi i(x)$ и называемых базовыми функциями. Форма $f(x, w)$ равна:

$$f(x, w) = \varphi(x)T\ w \qquad (1)$$

где w – весовой вектор или матрица $w = (w1, ..., wD)T$, а $\varphi = \varphi(1, ... ,D)T$.

· Логистическая регрессия – это аналитический метод, аналогичный линейной регрессии, который обнаруживает уравнение для прогнозирования результата для бинарной переменной Y из одной или нескольких переменных отклика, X. Переменные отклика могут быть категориальными или непрерывными, в отличие от линейной регрессии, поскольку модель не требует непрерывных данных. Для прогнозирования членства в группах и размещения конечного типа вместо метода наименьших квадратов используется итерационный метод максимального правдоподобия, использующий логарифмическое отношение шансов, а не вероятности [15].

4) Adaboost, известный как «Адаптивное повышение», представляет собой алгоритм метаобучения, который объединяет слабые классификаторы в уникальный сильный классификатор. AdaBoost позволяет классифицировать экземпляры, придавая больший вес сложным и меньший тем, которые уже обработаны хорошо. AdaBoost предназначен для решения задач классификации и регрессии [16].

3.3. Обнаружение аномальных писем

Инсайдеры могут нанести огромный ущерб в текущей ИТ-среде, и мы не можем полностью предотвратить ущерб от этого. Существующие средства и системы информационной безопасности зачастую нуждаются в усовершенствовании для смягчения возможных последствий инсайдерского воздействия. Это выражается в соответствующем запросе со стороны профильных специалистов в области информационных технологий и защиты информации, для которых исследования новых путей выявления аномальной активности особенно ценны [17, 18]. Для определения предела соответствия автоматизации информационной безопасности также следует изучить человеческий фактор. Для этого электронная почта отлично подходит в качестве среды для обнаружения инсайдерских угроз [19]. Наконец, чтобы получить представление о реальной применимости описанных методов и алгоритмов машинного обучения, приступим к экспериментальной части исследования.

4. Эксперименты и результаты

Реализована система на Python с Tensorflow в качестве бэкенда. Для разработки использовалась Anaconda IDE. Чтобы оценить возможные аномальные электронные письма, нужно выполнить пять шагов. На первом этапе данные собираются из репозитория НД. Этапы предварительной обработки включают в себя обнаружение

недостающих значений, удаление стоп-слов, стемминг и токенизацию. Затем следует преобразование данных, при котором текстовые данные преобразуются в векторную форму. После предварительной обработки для классификации электронных писем были применены алгоритмы машинного обучения. Эти шаги описаны поэтапно в последующих разделах, что сделано для удобства читателей и, в первую очередь, исследователей в рассматриваемой и смежных научных областях, в частности, для специалистов по Data Science и аналитиков данных. Разделение позволяет определить место и роль работы в общей последовательности действий для каждого из них в соответствии с их профилем. Привлечение внимания разного рода специалистов к рассматриваемой тематике особенно важно в том числе из соображений необходимости дальнейших исследований.

4.1. Очистка данных

Объединение всего набора данных и его анализ на предмет согласованности и любых несоответствий, требующих исправления. Кроме того, все строки, содержащие нулевые значения текста сообщения, удаляются или заполняются, как это предложено в определенной схеме.

4.2. Предварительная обработка данных

Набор данных состоит из файлов .csv, которые отправляются по электронной почте от разных пользователей. Все файлы электронной почты объединяются в один файл. В нем есть идентификатор пользователя, содержимое электронной почты, твиты и метка электронной почты в файле. После этого выполняется предварительная обработка набора данных, состоящая из следующих шагов.

1) Удаление стоп-слов – то есть слов, которые в одном предложении не имеют никакого эффекта. Вы можете легко игнорировать их, не теряя при этом контекста предложения.

2) Стемминг – метод минимизации слова до его основы, которая прикрепляется к суффиксам и префиксам или к корням слов, известный как лемма, называется стеммингом. Стемминг жизненно важен для обработки и понимания естественного языка.

3) Токенизация – это метод разделения текста на части, называемые токенами. Токены могут представлять собой отдельные слова, фразы или даже целые предложения. В методе токенизации можно отказаться от некоторых символов, например, знаков препинания. Токены обычно используются в качестве входных данных для таких процессов как векторизация.

4.3. Преобразование данных

Преобразование данных — это метод преобразования данных из одного формата в другой. Данные электронной почты состоят из текстовых данных, которые преобразуются в векторный формат, как показано на рисунке 2. После преобразования текстовых данных в векторную форму векторизация TF-IDF изменяет кластер необработанных каталогов на матрицу функций TF-IDF, как показано на рисунке 3.

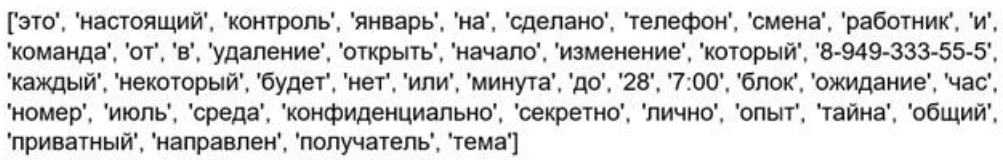

Рис. 2. Векторное представление текстовых данных

```
(0, 67767)    0.06937179449416819
(0, 47504)    0.3181875690631305
(0, 54789)    0.06792312041150975
(0, 38522)    0.09171483341147117
(0, 65942)    0.08369842513125694
(0, 73477)    0.14727563973076654
(0, 55895)    0.07271542548196704
(0, 55834)    0.082591831756838
(0, 74440)    0.03001431013948936
(0, 39721)    0.08716918875502584
(0, 31817)    0.09028669441540148
(0, 69267)    0.24174194244617922
(0, 51728)    0.032439071476644
(0, 52589)    0.05118235333859011
(0, 34813)    0.08022860520331387
```

Рис. 3. Кодирование данных через TF-IDF

4.4. Методы машинного обучения

Алгоритм для предложенного фреймворка на Python с тензорным потоком в бэкенде демонстрируется в Табл. 1.

Таблица 1

Алгоритм контролируемого обучения

Входные данные	Обучающие данные *D = x,y*
Выходные данные	Разделение обычных и вредоносных писем
Шаг 1	Предварительная обработка данных и сокращение сроков
Шаг 2	Формирование матрицы TF-IDF
Шаг 3	Изучение классификаторов AdaBoost, *k* ближайшего соседа, метода опорных векторов, логистической и линейной регрессии, алгоритм Байеса на основе набора данных
Возврат	*St(x) = h(AdaBoost)*

На наборе данных было запущено несколько отдельных классификаторов. А именно, AdaBoost, *k* ближайшего соседа, метод опорных векторов, логистическая регрессия, линейная регрессия и алгоритм Байеса. Одиночные классификаторы дают удовлетворительную производительность, с максимальным значением 0,983. Производительность модели измерялась с помощью таких показателей производительности как точность (Точн) и запоминание (Зап). Результаты, полученные с помощью моделей, представлены в Табл. 2 и на рисунке 4.

Здесь,

$$Точн = \frac{П + О}{П + О + ЛП + ЛО} \quad (3)$$

$$Зап = \frac{П}{П + ЛО} \quad (4)$$

где:

П = положительные (верно определённые нормальные эл. письма);

О = отрицательные (верно определённые вредоносные эл. письма);

ЛП = ложноположительные (вредоносные эл. письма, принятые за нормальные);

ЛО = ложноотрицательные (нормальные эл. письма, принятые за вредоносные).

Таблица 2

Результаты одиночных классификаторов на тестовом наборе

Классификатор	Точн (%)	Зап (%)	Матрица	
			Нормальные	Вредоносные
AdaBoost	98,3	98	495(П)	8(ЛП)
			9(ЛО)	488(О)
k ближайшего соседа	96	97	488(П)	24(ЛП)
			16(ЛО)	472(О)
Метод опорных векторов	95	95	478(П)	24(ЛП)
			16(ЛО)	472(О)
Логистическая регрессия	94,8	93	479(П)	27(ЛП)
			25(ЛО)	469(О)
Линейная регрессия	93,2	91	450(П)	18(ЛП)
			52(ЛО)	478(О)
Алгоритм Байеса	73,7	76	374(П)	166(ЛП)
			130(ЛО)	336(О)

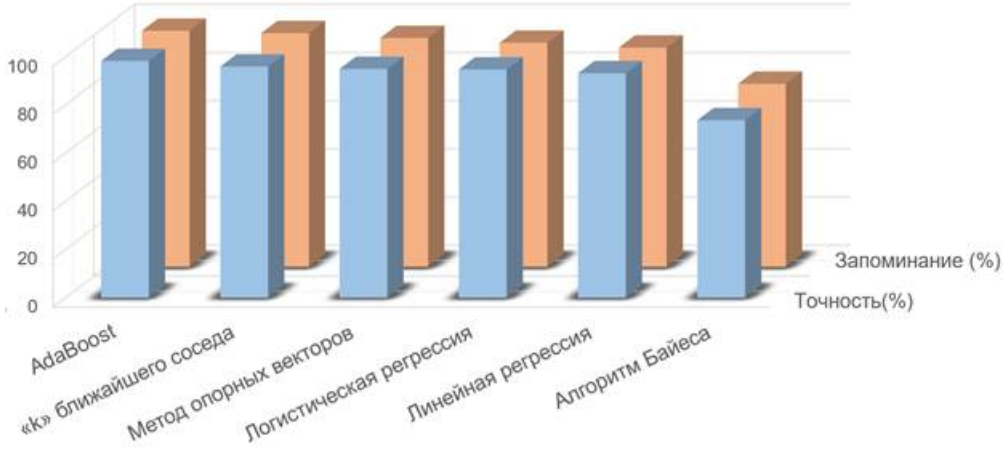

Рис 4. График, представляющий результаты алгоритмов обучения с учителем

Изучив полученные результаты, мы приходим к выводу, что среди предложенных алгоритмов, лучшим является метод "адаптивного повышения" (AdaBoost), так как позволяет получить наивысшее значение показателей точности и запоминания. Первый показатель характеризует возможность верной идентификации как нормальных, так и вредоносных объектов из всего их числа, демонстрируя реальную практическую применимость метода в рамках задачи поиска аномального поведения в информационных системах организаций. Второй показатель — это отсылка к возможностям машинного обучения, теоретически представленных в разделе 3.2. Только теперь это уже эмпирически полученная оценка их возможностей, она также может рассматриваться как косвенная характеристика глубины обученности самого алгоритма вкупе с его применимостью к подобной постановке задачи. Дословно это можно описать как возможность идентификации объектов, не содержащих аномалий, на основе их воспроизведения по памяти из числа нормальных объектов.

Стоит также упомянуть, что при комплексном использовании нескольких алгоритмов

одновременно результаты в виде характеристик точности и запоминания не суммируются и не возрастают, а результирующая эффективность совпадает с результатами метода, обладающего максимальной эффективностью среди выбранных в каждом конкретном случае. Так, например, комплекс, состоящий из классификаторов "метод опорных векторов" и "логистическая регрессия" показывают результативность равную 95% по обоим показателям, равно как и одиночный классификатор "метод опорных векторов".

5. Заключение

В статье исследованы методы идентификации аномальных электронных писем сотрудников. На основании экспериментов можно заключить, что достичь лучшей точности классификации почтовых сообщений возможно, используя комбинацию методов предварительной обработки, таких как TF-IDF и модель AdaBoost, которая правильно классифицирует 98,3% вредоносных электронных писем. Общедоступный набор данных, построенный на основе реалистичных примеров инсайдерских угроз, был использован для экспериментов с использованием различных технологий машинного обучения, результат которых превзошел существующие методы, применяемые к этому набору данных. Полученные результаты могут быть крайне полезны специалистам в области информационных технологий и защиты информации, в частности, аналитикам данных, специалистам по Data Science, научным исследователям в рассматриваемой и смежных областях, а также для технических специалистов силовых ведомств.

Сравнивая полученные значения точности идентификации инсайдерских угроз с результатами подходов, предложенных ранее зарубежными исследователями, обнаруживаем, что их удалось превзойти по ряду показателей. Так, например, точность выявления аномалий, полученная в работе Дюка Ли и др. [6], как уже было обозначено ранее в разделе 2, составила примерно 85%, тогда как нами достигнута точность равная 98,3%. Предложенная Мальвикой Сингх и др. [7] модель машинного обучения, состоящая из «сверхточных нейронных сетей», продемонстрировала точность порядка 90%, что также удалось превзойти при помощи пяти из шести опробованных классификаторов. Серьёзными оппонентами в этой сфере по праву могут считаться Диоп Абдулай и др. [5], объединившие методы машинного обучения и методы, основанные на процедурах линейной алгебры и параллельных вычислений, добившись при этом точности, в отдельных случаях достигающей 99%. Противопоставляя полученные результаты, следует отметить, что предложенный нами подход позволяет, пользуясь одними лишь методами контролируемого машинного обучения, получить показатели, уступающие менее чем на один процент, что колеблется в пределах допустимой погрешности при относительно небольшом массиве используемых входных данных. При этом продемонстрированный подход является привлекательной альтернативой для конечного разработчика программного решения, так как позволяет упростить его, ограничившись алгоритмами машинного обучения и не нагромождая программный код избыточными алгебраическими вычислениями.

Потенциальным направлением для будущих исследований является адаптация данной модели под конкретные технические средства и решения, широко известные и повсеместно применяемые в качестве основных компонентов системы информационной безопасности. Кроме того, не стоит забывать, что в рамках эксперимента задействован массив, ограничивающийся тысячей записей. Это ограничение продиктовано мощностями используемого на тестовом стенде оборудования. Большое значение имеют дальнейшие опыты в формате предложенной модели на более крупном наборе входных данных, приближенном к реальному объёму, который циркулирует в информационных

системах организаций, с использованием соответствующего программно-аппаратного обеспечения, ресурсов которого достаточно для проведения необходимых расчетов и эмуляций в большем количестве. Это позволит скорректировать процент точности обнаружения аномалий различными классификаторами, вероятно, даже выявить нецелесообразность применения некоторых из них. Но не только количественное, но и качественное усовершенствование описанных методов уместно в дальнейшем развитии затронутой проблематики. Так, если сделать акцент на более глубоком машинном обучении, то существует высокая вероятность увеличить эффективность предложенных методов, не прибегая при этом к помощи ресурсоёмкого оборудования, данное направление исследований в большей степени является прерогативой специалистов по Data Science.

Библиография

1. Шугаев В.А., Алексеенко С.П. Классификация инсайдерских угроз информации // Вестник воронежского института МВД России. 2020. № 2. С. 143-153.
2. Nicola d'Ambrosio, Gaetano Perrone, Simon Pietro Romano. Including insider threats into risk management through Bayesian threat graph // Computers & Security. 2023. No. 133. Pp. 1-21.
3. Karen Renaud, Merrill Warkentin, Ganna Pogrebna, Karl van der Schyff. VISTA: An Inclusive Insider Threat Taxonomy, with Mitigation Strategies // Information & Management. 2023. No. 60(8). Pp. 1-37.
4. Omar, S., Ngadi, A., and Jebur, H. H. Machine learning techniques for anomaly detection: an overview // International Journal of Computer Applications. 2013. No. 79(2).
5. Diop, A., Emad, N., Winter, T., Hilia, M. Design of an Ensemble Learning Behavior Anomaly Detection Framework // International Journal of Computer and Information Engineering. 2019. No. 13(10). Pp. 551-559.
6. D. C. Le, N. Zincir-Heywood and M. I. Heywood. Analyzing Data Granularity Levels for Insider Threat Detection Using Machine Learning // IEEE Transactions on Network and Service Management. 2020. No. 17(1). Pp. 30-44.
7. M. Singh, B. M. Mehtre and S. Sangeetha. User Behavior Profiling using Ensemble Approach for Insider Threat Detection // IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA), Hyderabad, India, 2019. P. 1-8.
8. Jiuming Lu and Raymond K. Wong. Insider Threat Detection with Long Short-Term Memory // Proceedings of the Australasian Computer Science Week Multiconference. New York: Association for Computing Machinery, 2019. P. 1-10.
9. Gavai, G. Sricharan, K. Gunning, D. Hanley, John Singhal, M. Rolleston, Robert. Supervised and Unsupervised methods to detect Insider Threat from Enterprise Social and Online Activity Data // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA). 2015. No. 6(4). Pp. 47-63.
10. Flavio Homoliak, Harill, Athul Toffalini, John Guarnizo, Ivan Castellanos, Juan Mondal, Soumik Ochoa, Mart´ın. The Wolf of SUTD (TWOS): A dataset of malicious insider threat behavior based on a gamified competition // Journal of Wireless Mobile Networks. 2018. No. 9(1). Pp. 54-85.
11. Kenyhercz, Michael W. and Passalacqua, Nicholas V. Missing data imputation methods and their performance with biodistance analyses // Biological Distance Analysis, 2016. P. 181-194.
12. Askari, Armin, Alexandre d'Aspremont, and Laurent El Ghaoui. Naive feature selection:

sparsity in naive bayes // International Conference on Artificial Intelligence and Statistics, 2020. P. 1813-1822.

13. Багаев И. В., Коломенская М. Д., Шатров А. В. Алгоритм наивного метода Байеса в задачах бинарной классификации на примере набора данных santander с платформы kaggle // Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: сб. ст. по материалам Четвертой всерос. науч.-практ. конф. Ч. I. / Перм. гос. нац. исслед. ун-т. Пермь, 2019. С. 32-36.
14. George A.F., Alan J. Lee. An overview on theory and algorithm of support vector machines // Journal of University of Electronic Science and Technology of China. 2012. No. 1 (40). Pp. 2-10.
15. Minjiang Fang, Dinh Tran Ngoc. Building a cross-border e-commerce talent training platform based on logistic regression model // The Journal of High Technology Management Research. 2023. No. 34(2). Pp. 1-12
16. Соколов В., Кузьминых Е., Гита Б. Аутентификация на основе мозговых волн с использованием слияния функций // Компьютеры и безопасность. 2023. № 129. С. 1-12.
17. Поляничко М.А. Базовая методика противодействия внутренним угрозам информационной безопасности // Вестник современных исследований. 2018. № 9.3 (24). С. 314-317.
18. Карпунина К.И. Проблемы обеспечения информационной безопасности российских предприятий в условиях кризиса // Экономическая безопасность общества, государства и личности: проблемы и направления обеспечения. Сборник статей по материалам IX научно-практической конференции / под общ. ред. Тактаровой С.В., Сергеева А.Ю. М.: Издательство «Перо», 2022. С. 210-213.
19. Arnau Erola, Ioannis Agrafiotis, Michael Goldsmith, Sadie Creese. Insider-threat detection: Lessons from deploying the CITD tool in three multinational organisations // Journal of Information Security and Applications. 2022. No. 67. Pp. 1-22.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Предмет исследования. Статья, исходя из названия, должна быть посвящена применению методов машинного обучения для противодействия инсайдерской угрозе информационной безопасности. Содержание статьи соответствует заявленной теме.

Методология исследования базируется на использовании методов анализа и синтеза данных. Ценно, что автор графически демонстрирует представленные результаты. Автор также утверждает, что «набор данных был собран на основе фактического взаимодействия пользователя с хост-машиной, который содержит как легитимные пользовательские данные, так и вредоносные инсайдерские экземпляры». Это говорит о глубоком погружении в содержание рассматриваемых вопросов, т.к. выводы автора должны строить на изучении реальных числовых данных.

Актуальность исследования вопросов, связанных с противодействием инсайдерской угрозе информационной безопасности, не вызывает сомнения, т.к. это отвечает национальным интересам Российской Федерации и вносит позитивный вклад в обеспечение технологического суверенитета Российской Федерации.

Научная новизна в представленном на рецензирование материале существует. В частности, она заключается с графическим представлением векторного представления текстовых данных. Рекомендуется также в тексте статьи указать потенциальных пользователей полученных результатов исследования.
Стиль, структура, содержание. Стиль изложения преимущественно научный, но отдельные суждения сделаны в разговорном жанре: например, автор говорит «Мы успешно справились с чрезмерной подгонкой в небольшом наборе данных, используя простые модели». Данный стиль для научной статьи не допускается. Структура статьи, в целом, выстроена грамотно. Наполнение содержания статьи сделано на высоком уровне в части первых структурных элементов, однако заключительная часть выполнена поверхностно. В частности, автор делает вывод о том, что «Таким образом, предложенный метод AdaBoost позволяет получить наивысшее значение точности.» На основании чего сделан такой вывод? Из текста статьи логически это не вытекает: необходимо обязательно сделать увязку вывода с изложенным текстом. Более того, обращает на себя внимание отсутствие перечня выявленных проблем при применении перечисленных выводов и вариантов устранения их. Также было бы интересно изучить возможность применения методов в комплексе. Возможно, у автора есть собственная методика, базирующаяся на использовании комплекса методов. В заключительной части статьи было бы также интересно обозначить потенциальные направления дальнейшего исследования.
Библиография. Ознакомление со списком литературы позволяет сделать вывод о том, что в нём содержится 15 наименований. При этом обращает на себя внимание тот факт, что 14 из 15 источников являются зарубежными. При доработке статьи автору рекомендуется увеличить количество используемых отечественных научных публикаций. Это позволит учесть современные российские тенденции в изучении вопросов применения методов машинного обучения для противодействия инсайдерской угрозе информационной безопасности.
Апелляция к оппонентам. Несмотря на наличие сформированного списка источников, какой-либо научной дискуссии по результатам проведённого исследования в тексте статьи не представлено. При этом, это позволило бы усилить уровень научной новизны и глубину погружения в поднимаемые вопросы.
Выводы, интерес читательской аудитории. С учётом всего вышеизложенного, статья требует корректировки (как минимум, в заключении), после проведения которой может быть решён вопрос о целесообразности её опубликования. Потенциально данная статья будет обладать высоким спросом со стороны широкого круга читателей: преподавателей, научных сотрудников, аналитиков и представителей органов государственной власти Российской Федерации и субъектов Российской Федерации.

Результаты процедуры повторного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Рецензируемая работа посвящена обеспечению кибербезопасности и противодействию инсайдерской угрозе информационной безопасности с применением методов машинного обучения для обнаружения аномальных электронных писем.
Методология исследования базируется на анализе наборов данных, содержащих примеры инсайдерских угроз, поступающих на электронную почту, использовании классических методов машинного обучения, применении алгоритмов обучения с учителем Adaboost (адаптивного повышения), Байеса, логистической и линейной

регрессии, метода опорных векторов.

Актуальность работы авторы связывают с тем, что предотвращение злонамеренных действий инсайдеров в организационной системе является важной задачей для кибербезопасности, а существующие средства и системы информационной безопасности зачастую нуждаются в усовершенствовании для смягчения возможных последствий инсайдерского воздействия.

Научная новизна рецензируемого исследования, по мнению рецензента состоит в выводах о том, что электронные письма являются важным атрибутом для обнаружения внутренних угроз, а также адаптации методов машинного обучения для идентификации аномальных электронных писем сотрудников и обеспечения кибербезопасности организаций.

В тексте статьи выделены следующие разделы: Введение, Существующие подходы, Методология исследования, Эксперименты и результаты, Заключение, Библиография.

В статье представлен краткий и ёмкий обзор имеющихся научных работ, посвященных рассматриваемым в публикации вопросам; отражены основные этапы проделанной авторами работы по сбору данных, их предварительной подготовке для анализа, в доступной форме изложены использованные платформы машинного обучения, алгоритмы обработки данных и языки программирования; показаны результаты обучения на тестовом наборе данных. Текст иллюстрирован 2 таблицами и 4 рисунками, изложение материала сопровождается 4 формулами.

Библиографический список включает 19 источников – научные публикации на русском и английском языках по рассматриваемой теме, на которые в тексте приведены адресные ссылки, что подтверждает наличие апелляции к оппонентам.

Для улучшения публикации авторам предлагается скорректировать заголовок статьи, акцентировав внимание не на машинном обучении, а на противодействии инсайдерской угрозе информационной безопасности – в таком варианте статья органично впишется в тематику журнала «Вопросы безопасности», поскольку в представленном варианте работа выглядит ближе к ИТ-сфере, нежели к сфере обеспечения безопасности. Например, возможна такая формулировка: «Противодействие инсайдерской угрозе информационной безопасности с применением методов машинного обучения». Также следует отметить, что рассмотренные в статье подходы ещё предстоит апробировать на фактических реальных данных, а не только на общедоступных их наборах.

Рецензируемый материал соответствует направлению журнала «Вопросы безопасности», отражает результаты проведенной авторами работы, содержит элементы научной новизны и практической значимости, может вызвать интерес у читателей, рекомендуется к опубликованию с учетом высказанных пожеланий.