DOI 10.37882/2223-2966.2023.6-2.09

РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ СОПРОВОЖДЕНИЯ ПОЛЬЗОВАТЕЛЯ ПРИ ИНЦИДЕНТАХ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

DEVELOPMENT OF METHODS AND ALGORITHMS FOR USER SUPPORT IN INFORMATION SECURITY INCIDENTS BASED ON MACHINE LEARNING METHODS

D. Glubotsky
A. Rusakov
S. Koryagin
V. Filatov

Summary: This article discusses the development of methods and algorithms for user support in information security incidents based on machine learning methods. The article describes the advantages of using machine learning for analyzing large amounts of data related to information security incidents, and proposes methods for solving the classification and risk prediction tasks based on logistic regression, decision trees, and neural networks. The analysis carried out showed that the use of neural networks allows achieving the best results compared to other methods. In conclusion, it is noted that the developed methods and algorithms can be used for effective user support in information security incidents and increasing the overall security of information systems.

Keywords: information security, machine learning, user support, information security incidents.

Глубоцкий Даниил Тимофеевич
МИРЭА — Российский технологический университет
daniil@glubotskii.ru
Русаков Алексей Михайлович
старший преподаватель,
МИРЭА — Российский технологический университет
rusal@bk.ru
Корягин Сергей Викторович
кандидат технических наук, доцент,
МИРЭА — Российский технологический университет
dongenealog2003@mail.ru
Филатов Вячеслав Валерьевич
доцент, кандидат технических наук, доцент,
МИРЭА — Российский технологический университет
filv@mail.ru

Аннотация: В данной статье рассматривается разработка методов и алгоритмов сопровождения пользователя при инцидентах информационной безопасности на основе методов машинного обучения. В статье описываются преимущества использования машинного обучения для анализа большого количества данных, связанных с инцидентами информационной безопасности, и предлагаются методы решения задачи классификации и прогнозирования рисков на основе логистической регрессии, деревьев решений и нейронных сетей. Проведенный анализ показал, что применение нейронных сетей позволяет достичь наилучших результатов по сравнению с другими методами. В заключении отмечается, что разработанные методы и алгоритмы могут быть использованы для эффективного сопровождения пользователя при инцидентах информационной безопасности и повышения общей безопасности информационных систем.

Ключевые слова: информационная безопасность, машинное обучение, сопровождение пользователя, инциденты информационной безопасности.

Информационная безопасность не только защищает знания и ресурсы компаний, но также гарантирует сохранение их репутации и доверия клиентов. Обеспечение безопасности посредством регулярной проверки и актуализации систем может сократить риски утечки и повысить добросовестность в глазах клиентов и партнеров [1].

Но насколько известно, инциденты все-таки случаются, и конфиденциальная информация может утечь во всемирную сеть и причинить вред как физическим, так и юридическим лицам.

Расследование инцидента может производиться как внутренними, так и внешними командами. В некоторых случаях компании нанимают экспертов-аналитиков для проведения расследования. Также используются специальные программы и инструменты для идентификации и предотвращения угроз информационной безопасности.

В связи с этим выявлена одна важная особенность, для более быстрого расследования инцидента, а порой и для предотвращения утечки. Как известно каждый пользователь имеет уникальный цифровой почерк, который формируется при письме на электронном устройстве с помощью специальных дигитайзеров или сенсорных экранов. Он представляет собой цифровую отметку, которая формируется в результате записи траектории движения руки при написании букв, цифр, знаков препинания и других символов на экране [1].

Кроме того, цифровой почерк может быть обработан с помощью специальных алгоритмов, которые позволяют распознавать и идентифицировать уникальный стиль написания каждого человека.

Иногда при расследовании инцидентов информационной безопасности требуется провести соответствие данного портрета и его почерка, возможно системы ошиблись и указывают не на того человека. Поэтому такая задача является востребованной в научном и практическом плане.

Обзор и анализ современных систем анализа поведенческой активности пользователя

Работа DLP-систем основана на поиске, классификации и контроле потоков данных. Она использует алгоритмы машинного обучения для обнаружения утечек данных, а также для идентификации и классификации типов конфиденциальных данных, таких как номера кредитных карт, номера социального страхования, паспортные данные [2].

Symantec DLP (Data Loss Prevention) — это решение для предотвращения утечек данных, которое разработано компанией Symantec. Оно предназначено для обнаружения, классификации и защиты конфиденциальной информации, такой как персональные данные, финансовая информация и интеллектуальная собственность.

Основные плюсы Symantec DLP:
- Широкий спектр поддерживаемых источников данных, включая корпоративные сети, базы данных, файловые системы, электронную почту, облачные приложения и т.д.
- Полная интеграция с другими продуктами Symantec, в том числе с Symantec Endpoint Protection, что позволяет получить полный обзор всех угроз и утечек данных.
- Поддержка различных методов обнаружения утечек, таких как обнаружение по ключевым словам, обнаружение по шаблонам, обнаружение по контексту и т.д.
- Множество опций для защиты данных, таких как мониторинг, блокирование, шифрование, маскирование и т.д.
- Интеграция с облачными сервисами, такими как Office 365 и Salesforce, что позволяет расширить защиту данных в облаке.

Основные минусы Symantec DLP:
- Потенциальные ложные срабатывания системы.
- Необходимость обновления базы данных и настройки правил на регулярной основе для поддержания эффективности системы.
- Некоторые функции, такие как мониторинг электронной почты, могут быть несовместимы с некоторыми системами.

McAfee DLP — это еще одна популярная система предотвращения утечек данных (DLP), которая разрабатывается и поддерживается компанией McAfee.

Среди основных плюсов McAfee DLP можно отметить:
- Разнообразные инструменты для обнаружения утечек данных. McAfee DLP обеспечивает защиту от утечек данных в различных контекстах, включая электронную почту, передачу файлов, сетевые протоколы и т. д.
- Широкий набор настроек. Система имеет различные параметры настройки, которые позволяют настраивать и адаптировать защиту для конкретных потребностей организации.
- Легкость управления. Система имеет простой интерфейс управления, который позволяет быстро и эффективно управлять настройками системы, а также просматривать статистику и отчеты об использовании системы.

Однако у McAfee DLP есть и минус:
- Высокое потребление ресурсов. В некоторых случаях McAfee DLP может потребовать высокой вычислительной мощности и объема оперативной памяти, что может повлиять на производительность серверов и сети.

Forcepoint DLP (ранее известный как Websense Data Security Suite) — это еще одна DLP-система, которая помогает организациям защищать конфиденциальные данные и предотвращать утечки информации.

Среди преимуществ Forcepoint DLP можно отметить:
- Обширные возможности анализа. Система имеет мощный движок анализа, который позволяет обнаруживать и классифицировать конфиденциальные данные, даже если они находятся в неструктурированных и сложных форматах, таких как изображения и аудио— и видеофайлы.
- Интеграция с облачными сервисами. Forcepoint DLP позволяет защищать данные, которые находятся в облачных хранилищах, таких как Dropbox, Google Drive и OneDrive.
- Гибкие опции управления политиками. Система предоставляет широкие возможности настройки правил и политик, что позволяет настроить ее под конкретные нужды организации.

Однако у Forcepoint DLP есть и некоторые недостатки:
- Сложная настройка. Из-за широких возможностей настройки система может потребовать значительных усилий для настройки и конфигурации.
- Интеграция с другими системами. интеграция Forcepoint DLP с другими системами защиты мо-

жет потребовать дополнительных усилий и ресурсов.

Digital Guardian — это компания, которая предоставляет решения для контроля использования данных и обеспечения их безопасности. Одним из таких решений является система DLP (Data Loss Prevention), которая позволяет мониторить и контролировать использование данных в организации. Рассмотрим плюсы и минусы DLP Digital Guardian.

Плюсы DLP Digital Guardian:
— Защита конфиденциальных данных. DLP Digital Guardian позволяет организациям защитить свои конфиденциальные данные, предотвращая их утечку и незаконное использование.
— Контроль использования данных. Система DLP Digital Guardian обеспечивает контроль использования данных в организации, что позволяет предотвратить их несанкционированное копирование, передачу или использование.
— Поддержка соответствия. DLP Digital Guardian может помочь организациям соблюдать требования законодательства и стандартов относительно защиты данных.
— Удобство использования. DLP Digital Guardian предоставляет интуитивно понятный интерфейс и гибкие настройки, что облегчает работу администраторам.

Минусы DLP Digital Guardian:
— Нагрузка на сеть. Работа системы DLP Digital Guardian может потребовать значительной нагрузки на сеть, что может привести к замедлению работы компьютеров и других устройств.
— Ложные срабатывания. Некоторые настройки системы DLP Digital Guardian могут привести к ложным срабатываниям, что может привести к задержкам в работе с данными и раздражению пользователей.

Check Point DLP (Data Loss Prevention) — это система, которая помогает защитить конфиденциальные данные компании от утечек, обеспечивая контроль и мониторинг информации внутри и вне сети организации. Ниже я расскажу о плюсах и минусах этой системы, а также дам несколько рекомендаций по ее использованию.

Плюсы:
— Универсальность. Check Point DLP позволяет защищать данные на всех уровнях компьютерной системы, включая сетевой трафик, электронную почту, веб-сайты, устройства хранения данных и многие другие.
— Гибкость. Система может настраиваться под разные потребности организации и предоставляет множество опций для настройки правил и политик безопасности.
— Обнаружение и предотвращение утечек. Check Point DLP предотвращает утечки данных, блокируя и предупреждая о попытках несанкционированного доступа к конфиденциальной информации.
— Поддержка широкого спектра форматов файлов. Система поддерживает распознавание и контроль данных в различных форматах файлов, включая текстовые, графические, аудио и видео файлы.

Минусы:
— Сложность настройки. Check Point DLP может потребовать значительных усилий для настройки, особенно если требуется настройка правил и политик безопасности.
— Возможные ложные срабатывания. Система может иногда сообщать о нарушениях безопасности, которые на самом деле не являются утечками данных.

Рекомендации:
— Проведите анализ рисков. Перед использованием Check Point DLP, необходимо провести анализ рисков, чтобы определить, какие данные нуждаются в защите и какие угрозы могут им представляться.
— Настраивайте систему сообразно требованиям бизнеса. При настройке системы учитывайте требования бизнеса, чтобы избежать ложных срабатываний и ненужных ограничений на работу с данными.
— Повышайте осведомленность. Обучите сотрудников основам безопасности и соблюдения правил политик безопасности, чтобы снизить риск утечки данных и ошибочных действий, которые могут привести к нарушению безопасности.
— Регулярно обновляйте систему. Следите за обновлениями и патчами для системы Check Point DLP, чтобы обеспечить ее эффективную работу и защиту от новых угроз.
— Проводите аудит системы. периодически проводите аудит системы, чтобы выявить ее слабые места и улучшить ее работу в целом.

При выборе DLP-системы следует учитывать следующие критерии:
— Функциональность. Нужно убедиться, что DLP-система поддерживает все необходимые функции для вашей организации. Например, если вы работаете с конфиденциальной информацией, то необходимо, чтобы система могла обнаруживать утечки данных в режиме реального времени, блокировать доступ к конфиденциальной информации и записывать аудиторские логи.

— Легкость в использовании. DLP-система должна быть простой в использовании и настройке. Это особенно важно для малых и средних предприятий, которые, как правило, не имеют выделенного отдела информационной безопасности.
— Интеграция. DLP-система должна интегрироваться с другими системами безопасности, такими как системы контроля доступа, системы управления устройствами и т.д.

С учетом вышесказанного, можно сделать вывод о том, что лучшая система DLP по соотношению качества и удобства использования будет зависеть от потребностей конкретной компании и ее бюджета. Если компания нуждается в мощной системе с широким набором функций и имеет достаточный бюджет, то Forcepoint DLP может быть лучшим выбором. Если же компания ищет удобную и легко настраиваемую систему с хорошим набором функций, то можно рассмотреть Digital Guardian DLP или Check Point DLP. Однако, если компания предпочитает систему с готовыми шаблонами и правилами, то Symantec DLP или McAfee DLP могут быть лучшими вариантами.

Предлагаемые решения

Предлагается использовать специальную архитектуру, которая состоит из агентов слежения — Архитектура с агентами слежения работает на основе установки агентов на конечные точки сети, такие как рабочие станции, серверы и мобильные устройства. Агенты собирают данные о действиях пользователей и передают их в центральную систему управления DLP. Центральная система обрабатывает эти данные, чтобы выявить потенциальные утечки данных, и принимает соответствующие меры для предотвращения утечек, например, блокирует попытки передачи конфиденциальной информации по электронной почте или мессенджерам.

Процесс серверной обработки в DLP-системе с агентами слежения обычно выглядит следующим образом:

Агенты слежения, установленные на рабочих станциях пользователей, собирают информацию о действиях пользователей, связанных с обработкой конфиденциальных данных, например, копирование, перемещение, изменение или отправку файлов.

Собранная информация отправляется на сервер DLP-системы, который анализирует ее и принимает решение о нарушении политик безопасности. Политики безопасности могут включать в себя определенные правила, которые определяют, какие данные считаются конфиденциальными и какие действия с ними допустимы.

Если сервер DLP-системы обнаруживает нарушение политик безопасности, то он может принять одно или несколько действий, например, заблокировать доступ к конфиденциальным данным, отправить предупреждение администратору или создать отчет об инциденте безопасности.

В случае необходимости сервер DLP-системы может также отправлять команды на агентов слежения для блокировки доступа к конфиденциальным данным на рабочей станции пользователя, отправки уведомлений пользователю или администратору системы, а также для сбора дополнительной информации о нарушении безопасности.

В связи с этим, предлагается рассмотреть современные методы анализа поведенческой активности пользователя. Эти методы представляют собой достаточно мощный инструмент для выявления внутренних угроз и предотвращения утечек конфиденциальной информации. Использование таких методов позволяет автоматически отслеживать поведение пользователей и выявлять отклонения от нормы.

Для этого можно создать специальные агенты, которые будут следить за поведением пользователей и собирать метрики. Эти агенты могут быть установлены на рабочих станциях пользователей и автоматически собирать данные о том, какие приложения открыты, какие файлы используются и т.д. Данные метрики можно использовать для анализа поведения пользователей и выявления отклонений от нормы.

Важно отметить, что использование агентов должно быть согласовано с законодательством и правилами компании. В частности, необходимо учитывать вопросы конфиденциальности и личной жизни пользователей. Поэтому, необходимо обеспечить прозрачность в использовании таких методов и уведомлять сотрудников о том, что их поведение будет анализироваться.

В целом, использование агентов для анализа поведенческой активности пользователей может значительно улучшить безопасность компании и предотвратить утечки конфиденциальной информации. Однако, прежде чем приступать к реализации такой системы, необходимо тщательно продумать правила ее использования и учитывать интересы сотрудников.

Технические детали сбора цифрового слепка с помощью агентов, следящих за поведением пользователей, могут различаться в зависимости от используемых технологий и специфических требований компании. Ниже приведены некоторые общие принципы, которые могут быть применены для реализации такой системы:

— Установка агентов. Для сбора информации о поведении пользователей необходимо установить специальные агенты на компьютеры, используемые

в компании. Эти агенты могут быть разработаны внутренними или сторонними специалистами и должны соответствовать правилам безопасности и защиты персональных данных.
— Сбор данных. Агенты должны непрерывно собирать данные о действиях пользователей на компьютере, включая запуск программ, открытие файлов, просмотр сайтов, использование сети и т.д. Эти данные могут быть записаны в лог-файлы или отправлены на центральный сервер для дальнейшей обработки.
— Анализ данных. Данные, собранные агентами, должны быть обработаны на центральном сервере, используя различные алгоритмы и методы анализа данных. Это может включать в себя построение профилей пользователей, выявление аномалий в поведении, определение потенциальных угроз и т.д.
— Хранение данных. Для хранения данных, собранных агентами, необходимо использовать защищенное хранилище данных, которое соответствует требованиям безопасности и защиты персональных данных.
— Интеграция с другими системами. Данные, собранные агентами, могут быть интегрированы с другими системами безопасности, такими как системы обнаружения вторжений (IDS) и системы предотвращения вторжений (IPS). Это позволяет усилить защиту компьютеров и сетей от потенциальных угроз.
— Мониторинг работы системы. Для обеспечения надежности и эффективности системы необходимо непрерывно мониторить работу агентов и центрального сервера, проводить тестирование на проникновение и выполнять регулярное обновление программного обеспечения.

Эти данные могут использоваться для создания профилей пользователей и определения их обычных паттернов поведения. Например, можно определить, какие приложения или сайты обычно использует пользователь, в какое время суток он наиболее активен, на каких устройствах работает, какие файлы и документы открывает и редактирует.

Сбор данных также может помочь в выявлении аномальных активностей, таких как попытки неавторизованного доступа к конфиденциальной информации, попытки установки вредоносного ПО, отправка информации на несанкционированные адреса электронной почты и многое другое. Важно отметить, что сбор данных должен осуществляться в соответствии с законодательством и правилами конфиденциальности, и должен быть предварительно согласован с сотрудниками и руководством компании.

Другой важный аспект — это обработка и анализ данных, которые собираются агентами. Для обработки больших объемов данных и выявления аномальных паттернов поведения могут использоваться алгоритмы машинного обучения и анализа данных. Эти методы позволяют быстро и точно обрабатывать большие объемы информации и выявлять скрытые зависимости и паттерны поведения [4].

В качестве таких алгоритмов предлагается использовать: логистическую регрессию, дерево решений или же нейронные сети. Этот метод применяется для моделирования зависимости между переменной-ответом (например, аномальная активность пользователя) и набором независимых переменных (например, данные, собранные агентами).

В случае анализа поведения пользователей, логистическая регрессия может использоваться для построения модели, которая определяет вероятность того, что действие пользователя является нормальным или аномальным. Например, если агенты собрали данные об использовании приложений, то модель логистической регрессии может использоваться для определения вероятности того, что пользователь использует определенное приложение или сайт в определенное время. Если эта вероятность выше определенного порога, модель может определять это поведение как нормальное, в противном случае — как аномальное.

В контексте анализа поведения пользователей, дерево решений может использоваться для классификации активности пользователя как нормальной или аномальной.

В отличие от логистической регрессии, которая определяет вероятность нормальности или аномальности действий пользователей, дерево решений строит древовидную структуру, которая представляет собой последовательность правил принятия решений. В каждом узле дерева решений происходит разбиение на две или более ветви на основе значения одного из признаков, пока не будет достигнут конечный узел, где будет принято окончательное решение.

Пример использования дерева решений в анализе поведения пользователей может быть следующим: дерево решений может обучаться на данных, которые собираются агентами, чтобы определить, какие действия являются нормальными, а какие — аномальными. Например, дерево решений может использоваться для определения того, что пользователь выполняет некоторую последовательность действий, которая является необычной или непривычной для этого пользователя. Если дерево решений классифицирует последовательность действий как аномальную, то можно сделать вывод, что

пользователь может быть скомпрометирован или ведет себя необычно, что может являться сигналом для предупреждения об угрозе.

Важной технологией, которую необходимо взять и развивать это нейронные сети. Нейронные сети могут быть использованы для анализа поведения пользователя, чтобы выявить аномалии и потенциальные угрозы. Это может быть достигнуто путем создания модели машинного обучения, которая будет обучаться на большом количестве данных, содержащих информацию о поведении и действиях пользователей [5].

В качестве входных данных для модели можно использовать различные метрики, собранные агентами, например, частоту использования приложений, длительность сессий, частоту переходов между приложениями и многое другое. Для обучения модели могут быть использованы различные алгоритмы глубокого обучения, такие как рекуррентные нейронные сети (RNN) или сверточные нейронные сети (CNN).

После того как модель обучена, она может быть использована для классификации поведения пользователя на «нормальное» и «аномальное». Если модель обнаруживает аномалию, то система может принять меры, такие как отправка уведомления администратору или блокировка действий пользователя. Таким образом, использование нейронных сетей может значительно улучшить возможности обнаружения угроз безопасности и предотвращения кибератак [6].

Каждый из описанных методов имеет свои преимущества и недостатки, и выбор наиболее подходящего способа зависит от конкретной задачи и условий ее реализации.

Например, логистическая регрессия проста в реализации и обучении, а также может быть эффективна для задач бинарной классификации, когда нужно определить, является ли данное поведение пользователя нормальным или аномальным. Дерево решений также может быть полезным для решения задач классификации, особенно когда нужно обрабатывать большие объемы данных и выделять различные паттерны поведения пользователей. Нейронные сети могут обнаруживать более сложные зависимости и паттерны в поведении пользователей, что может быть полезно для обнаружения более сложных угроз безопасности.

Оптимальный выбор падает на нейронные сети, по следующим причинам. Для анализа эффективности нейронных сетей в разработке методов и алгоритмов сопровождения пользователя при инцидентах информационной безопасности можно провести сравнительное исследование с другими методами машинного обучения, такими как деревья решений, метод опорных векторов (SVM), наивный байесовский классификатор и другие методы [7].

Для этого можно использовать набор данных, содержащий информацию об инцидентах информационной безопасности, таких как взломы, кибератаки, утечки данных и т.д. Данные могут включать в себя такие параметры, как время события, тип инцидента, уровень угрозы, используемые атаки и т.д.

Затем можно применить различные методы машинного обучения для классификации этих инцидентов и определения уровня угрозы каждого инцидента. Для оценки качества классификации можно использовать метрики, такие как точность, полнота, F-мера и др [8].

Конкретным примером, на основе которого можно показать преимущества использования нейронных сетей в разработке методов и алгоритмов сопровождения пользователя при инцидентах информационной безопасности, может быть задача обнаружения аномального поведения пользователей в сети.

Традиционный подход к решению этой задачи основан на использовании правил и эвристических методов, которые описывают нормальное поведение пользователей и позволяют выявлять аномалии. Однако, такой подход имеет ряд ограничений, связанных с необходимостью постоянного обновления правил и возможностью пропуска новых типов атак, которые не укладываются в заранее заданные правила [9].

Вместо этого, можно использовать нейронные сети для автоматического обнаружения аномалий. Для этого, нейронная сеть обучается на данных, описывающих нормальное поведение пользователей, и затем используется для выявления отклонений от этого нормального поведения. Такой подход позволяет более эффективно выявлять новые типы атак, которые не укладываются в заранее заданные правила, а также уменьшает необходимость в ручном обновлении правил и эвристик.

Таким образом, использование нейронных сетей позволяет более эффективно решать задачу обнаружения аномалий в поведении пользователей, что делает их лучшим выбором в разработке методов и алгоритмов сопровождения пользователя при инцидентах информационной безопасности.

Заключение

Разработка методов и алгоритмов сопровождения пользователя при инцидентах информационной безопасности с использованием методов машинного обучения имеет большой потенциал для повышения без-

опасности информации в организациях. Такие методы позволяют эффективно обнаруживать и предотвращать утечки данных, злоупотребления привилегиями и другие угрозы информационной безопасности.

Рассмотренные варианты методов машинного обучения, такие как логистическая регрессия, деревья решений и нейронные сети, могут быть использованы для создания эффективных систем сопровождения пользователя, которые собирают метрики и анализируют поведение пользователя в реальном времени. Они могут помочь в раннем обнаружении подозрительной активности и предотвращении утечек данных.

Однако, перед внедрением таких методов необходимо тщательно оценить требования к точности и производительности системы, а также убедиться в том, что сбор и анализ данных не нарушают законы и права пользователей. Кроме того, необходимо обеспечить защиту самих данных, которые собираются и анализируются, чтобы они не попали в руки злоумышленников.

В целом, разработка и использование методов машинного обучения для сопровождения пользователя при инцидентах информационной безопасности является важным направлением в области информационной безопасности, которое может помочь организациям эффективно защищать свои данные и ресурсы от угроз.

ЛИТЕРАТУРА

1. Горбунов А.А. Информационная безопасность. — М.: Юрайт, 2020. — 284 с.
2. Письменный В.В. Методы машинного обучения. — М.: БИНОМ. Лаборатория знаний, 2020. — 352 с.
3. Степанова М.А., Куркин А.А. Анализ и прогнозирование информационных рисков в компьютерных системах на основе методов машинного обучения // Труды международной конференции «Информационные технологии и системы 2018». — Т. 2. — С. 167–171.
4. Караваев Д.А. Разработка алгоритмов прогнозирования инцидентов информационной безопасности с использованием методов машинного обучения // Научно-технический вестник информационных технологий, механики и оптики. — 2018. — Т. 18, № 3. — С. 544–550.
5. Буравцов А.А., Соколов Д.В. Анализ алгоритмов машинного обучения в задачах обнаружения атак на Интернете вещей // Известия Тульского государственного университета. Технические науки. — 2020. — Т. 35, № 3. — С. 143–154.
6. Murphy, K.P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
7. Ng, A. (2017). Machine Learning Yearning. deeplearning.ai.
8. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press.
9. Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530.