УДК 004.85 **DOI: 10.18413/2518-1092-2026-11-1-0-5**

Потиенко Д.А.
Газизов А.Р.
Легонько О.Л.

ЭКСПЕРТНАЯ СИСТЕМА АНАЛИЗА СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

ФГБОУ ВО «Донской государственный технический университет» (ДГТУ)
пл. Гагарина, 1, г. Ростов-на-Дону, 344002, Россия

e-mail: potienkodaniil@gmail.com, agazizov@donstu.ru, olga_cvetkova@mail.ru

Аннотация
Статья посвящена разработке структуре гибридной экспертной системы для анализа событий информационной безопасности, направленной на автоматизацию выявления угроз в сетевом трафике. В условиях роста объема данных и сложности атак традиционные методы анализа недостаточно эффективны, что обусловливает необходимость интеграции машинного обучения и экспертных правил. Цель исследования — создание модульной архитектуры системы, включающей четыре компонента: сбор данных (Apache Kafka), предварительную обработку (Apache Flink), анализ и классификацию (Random Forest с постобработкой правилами) и ведение журналов (Elastic Stack). Прототип на Python протестирован на датасете UNSW-NB15, демонстрируя точность бинарной классификации 0,890 и мультиклассовой классификации 0,781. Гибридный подход повышает показатель recall для выбранных классов атак (Analysis, Backdoor, DoS) на 19–100% при снижении общей точности на 1,2%, обеспечивая интерпретируемость решений. В заключении предлагаются направления развития, включая оптимизацию правил через обучение с подкреплением, интеграцию искусственных нейронных сетей LSTM и автоматическое обновление базы знаний.
Ключевые слова: экспертная система; технологии искусственного интеллекта; информационная безопасность; защита информации; угроза информационной безопасности
Для цитирования: Потиенко Д.А., Газизов А.Р., Легонько О.Л. Экспертная система анализа событий информационной безопасности // Научный результат. Информационные технологии. – Т.11, №1, 2026. – С. 40-53. DOI: 10.18413/2518-1092-2026-11-1-0-5

Potienko D.A.
Gazizov A.R.
Legonko O.L.

EXPERT SYSTEM FOR INFORMATION SECURITY EVENT ANALYSIS

Don State Technical University
1 Gagarin square, Rostov-on-Don, 344003, Russia

e-mail: potienkodaniil@gmail.com, agazizov@donstu.ru, olga_cvetkova@mail.ru

Abstract
This article explores the development of a hybrid expert system for analyzing information security events, aimed at automating threat detection in network traffic. Given the growing volume of data and the complexity of attacks, traditional analysis methods are ineffective, necessitating the integration of machine learning and expert rules. The goal of the study is to create a modular system architecture comprising four components: data collection (Apache Kafka), preprocessing (Apache Flink), analysis and classification (Random Forest with rule-based postprocessing), and logging (Elastic Stack). A Python prototype was tested on the UNSW-NB15 dataset, demonstrating a binary classification accuracy of 0,890 and a multiclass classification accuracy of 0,781. The hybrid approach increases recall for selected attack classes (Analysis, Backdoor, DoS) by 19–100% while reducing overall accuracy by 1,2%, ensuring the interpretability of solutions. The conclusion suggests future directions, including rule optimization through reinforcement learning, integration of LSTM artificial neural networks, and automatic knowledge base updating.
Keywords: expert system; artificial intelligence technologies; information security; information protection; information security threat

For citation: Potienko D.A., Gazizov A.R., Legonko O.L. Expert System for Information Security Event Analysis // Research result. Information technologies. – T.11, №1, 2026. – P. 40-53. DOI: 10.18413/2518-1092-2026-11-1-0-5

ВВЕДЕНИЕ

В современных условиях интенсивного развития информационных технологий постоянно увеличивается объем и сложность данных, которые обрабатываются в системах безопасности. Это обстоятельство существенным образом влияет на скорость анализа событий и инцидентов, регистрируемых в информационной инфраструктуре предприятий. Традиционные методы мониторинга и анализа, основанные на вручную составляемых правилах и экспертных оценках, не позволяют своевременно выявлять угрозы, связанные со сложными атаками и аномальными поведениями, так как такие угрозы часто маскируются под нормальную активность, что требует автоматизированных и адаптивных методов обнаружения. В связи с этим возникает необходимость автоматизации процессов анализа информации о безопасности – создания интеллектуальных систем, способных оперативно и точно обрабатывать большие объемы данных, выявлять потенциальные угрозы и предоставлять аналитические отчеты специалистам.

Актуальность рассматриваемой темы обуславливается необходимостью обеспечения и постоянного поддержания требуемых уровней защищенности информации. Внедрение методов искусственного интеллекта и машинного обучения позволяет улучшить качество анализа, снизить нагрузку на специалистов и сократить время реакции на инциденты, как показано в обзоре методов машинного обучения в системах обнаружения сетевых вторжений [1]. Однако реализация таких систем сопряжена с рядом технических и организационных задач, включая интеграцию различных источников данных, разработку гибридных моделей анализа и обеспечения объяснимых решений.

В научном исследовании [2] анализируются различные подходы к повышению эффективности систем обнаружения вторжений (IDS), включая методы машинного обучения, байесовские алгоритмы, метаэвристики, роевой интеллект и марковские нейронные сети, а также одиночные, гибридные и ансамблевые алгоритмы классификации, оцениваемые на множестве наборов данных. Авторы сравнивает метрики результатов, недостатки и используемые датасеты, предлагая будущие направления исследований для улучшения систем безопасности.

В обзорной статье [3] представлен обзор литературы по кибераналитике в поддержку методов обнаружения вторжений, основанных на машинном обучении (МО) и интеллектуальном анализе данных (ИИД). Авторы представили краткое руководство по методам МО/ИИД, рассмотрели проведенные исследования, а также выполнили классификацию работ на основе количества цитирований и значимости метода.

В работе [4] авторы указывают на тот факт, что обнаружение сетевых вторжений на основе аномалий играет ключевую роль в защите сетей от вредоносной активности, системы IDS стремятся к низким ложным срабатываниям и высокому обнаружению. Однако классификационные техники неэффективны для неизвестных атак. В статье выполняется разработка универсального метаэвристического подхода для обнаружения как известных, так и неизвестных атак с высоким уровнем обнаружения и низким ложным срабатыванием путем эффективной оптимизации признаков.

В сетевых пакетах может присутствовать атака, предоставляющая хакеру доступ к конфиденциальной информации компьютера, поэтому в научной работе [5] была создана модель обнаружения сетевых вторжений на основе классификатора Random Forest для предсказания безопасности пакетов с максимальной точностью. Для тестирования модели в реальном времени разработан анализатор пакетов, который преобразует сетевые данные в признаки и проверяет их на легитимность.

В статье [6] предложен подход к обнаружению сетевых вторжений (NIDS) с использованием классификатора случайного леса на датасете CICIDS-2017, включающий предварительную обработку данных, отбор 26 ключевых признаков и оптимизацию весов классов для преодоления ограничений традиционных сигнатурных методов. Результаты демонстрируют высокую

эффективность модели с 99,8% взвешенной F1-оценки и 93,31% макро-F1-оценки, подчеркивая потенциал машинного обучения в борьбе с эволюционирующими киберугрозами.

В исследовании [7] разработана система обнаружения сетевых вторжений с использованием различных классификаторов машинного обучения на датасете KDD99, включая предварительную обработку данных с анализом корреляций и реализацию моделей, таких как наивный байесовский алгоритм, дерево решений, случайный лес, SVM, логистическая регрессия и ансамблевое голосование. Анализ показал, что наивный байесовский классификатор имеет самую низкую точность, в то время как случайный лес демонстрирует наивысшую точность и лучшие результаты по метрикам, таким как точность, полнота, f1-критерий и время обучения.

В работе [8] предложена система обнаружения вторжений на основе машинного обучения с новой стратегией выбора признаков IV-RFE, которая учитывает относительную дисперсию и весовой фактор в сочетании с рекурсивным исключением признаков, чтобы решить проблемы низкой точности и высокой размерностью данных в IDS. В отличие от предыдущих исследований, фокусирующихся только на выборе признаков, данный подход также обеспечивает стабильность набора признаков.

Для обучения моделей авторы работы [9] подготовили сбалансированную выборку с признаками нормального и аномального трафика, выбрали пять алгоритмов машинного обучения и провели ряд тестирований. По результатам экспериментов был отобран классификатор случайного леса, показавший наилучшие результаты.

В научной статье [10] на основании проведенного анализа авторы разработали рекомендации по реализации систем выявления внутренних угроз с помощью алгоритмов машинного обучения.

В работе [11] представлен анализ ключевых задач информационной безопасности, с учетом применения искусственного интеллекта: обнаружение атак, вредоносных программ, модификаций данных, утечек информации, оценку рисков и повышение надежности компьютерных систем и сетей.

В статье [12] представлен модифицированный алгоритм отрицательного отбора и проведен вычислительный эксперимент с иммунной системой, обнаруживающей сетевые вторжения, демонстрирующий ответную защитную реакцию системы при обнаружении аномального объекта.

Целью данной статьи является разработка архитектуры и концептуальной модели экспертной системы анализа событий информационной безопасности, комбинирующей автоматический сбор данных, их обработку, а также интеграцию правил и моделей машинного обучения для повышения точности выявления угроз.

В рамках исследования поставлены задачи: разработать структуру модульной экспертной системы с четырьмя ключевыми модулями (сбор данных, предварительная обработка, анализ и классификация, ведение журналов и отчетность); оценить ее работоспособность на прототипе с использованием исследовательского набора данных UNSW-NB15; и провести экспериментальную валидацию гибридной классификации с постобработкой для редких классов атак. В качестве основных методов использованы алгоритмы машинного обучения (Random Forest для бинарной и мультиклассовой классификации) и экспертные правила на основе статистического анализа признаков (dload, dur, sbytes и др.).

ОПИСАНИЕ АРХИТЕКТУРЫ И КОМПОНЕНТОВ ЭКСПЕРТНОЙ СИСТЕМЫ АНАЛИЗА СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Целью разработки экспертной системы является создание автоматизированной системы, способной эффективно анализировать возникающие события в информационной инфраструктуре предприятия, связанные с сетевой активностью и интернет-трафиком, с целью обнаружения потенциальных угроз, атак и аномалий, а также классификации типов событий. Она позволит специалистам службы безопасности своевременно выявлять подозрительные активности, работать с журналами событий, а также формировать отчеты для принятия решений по защите информационных систем. Система предназначена для автоматического сбора и предварительной

обработки данных о безопасности, анализа и классификации событий с использованием сочетания базы знаний и моделей машинного обучения, а также для автоматического ведения журналов и формирования отчетов.

Разрабатываемая экспертная система представляет собой многоуровневую архитектуру, состоящую из четырех модулей, каждый из которых выполняет специализированные задачи в цепочке обработки данных. В таблице 1 представлено описание модулей проектируемой экспертной системы с указанием решаемых задач и методов реализации. Для понимания движения информационных потоков в проектируемой экспертной системе в таблице также показаны входные и выходные сигналы для каждого модуля, исходя из специфики решаемых задач.

Таблица

Компоненты и методы системы анализа событий информационной безопасности

Table

Components and methods of the information security event analysis system

№	Модуль	Задачи	Методы реализации	Входные / Выходные сигналы
1	Модуль сбора данных	Автоматический сбор информации из различных источников	Сбор данных в реальном времени с помощью системы Apache Kafka	Данные из журналов событий, систем обнаружения вторжений (IDS), сетевого трафика, систем управления безопасностью / Необработанные данные
2	Модуль предварительной обработки данных	Фильтрация, стандартизация данных для устранения шумов и ошибок, приведения данных к единому формату	Обработка данных в реальном времени с помощью системы Apache Flink: 1. Фильтрация на основе регулярных выражений и фильтров по IP-адресам 2. Стандартизация признаков с помощью метода Min-Max масштабирования (Scaling)	Необработанные данные / Стандартизированные данные (нормализованные числовые векторы и очищенные текстовые поля)
3	Модуль анализа и классификации событий	Выявление подозрительных событий, определение их класса (классификация)	1. Бинарная классификация с помощью алгоритма машинного обучения (метод случайного леса, Random Forest). Цель — разделение набора событий на нормальные события и атаки 2. Мультиклассовая классификация с помощью алгоритма машинного обучения (Random Forest). Цель — определение классов атак (исходными данными является массив атак, полученный на этапе бинарной классификации)	Стандартизированные данные / Классификации событий с указанием класса атак

№	Модуль	Задачи	Методы реализации	Входные / Выходные сигналы
			3. Классификация событий с помощью базы знаний (правил). Цель – уточнение предсказаний классов атак для событий с низкой уверенностью модели, чтобы повысить полноту обнаружения атак без полного переобучения модели	
4	Модуль ведения журналов и отчетности	Автоматическое ведение журнала и структурирование инцидентов для последующего анализа, аудита и подготовки отчетности, предоставление аналитической информации специалистам службы безопасности для принятия решений	Работа с информацией с помощью системы Elastic Stack (Elasticsearch для хранения и поиска информации, Kibana для визуализации) [13]	Классификации текущих событий с указанием класса атак, исторические статистические данные о событиях / Конечный выход системы для специалистов службы безопасности (структурированные отчеты, визуализации и аналитическая информация для аудита, принятия решений)

Предлагаемая модульная структура экспертной системы обеспечивает гибкость и масштабируемость системы, позволяя легко добавлять или обновлять отдельные компоненты без изменения всей архитектуры. Кроме того, модульность снижает сложность интеграции с внешними системами и технологиями, а также упрощает сопровождение и развитие системы в будущем.

В проектируемой экспертной системе в Модуле анализа и классификации событий предлагается использование гибридной архитектуры – сочетание правил и моделей машинного обучения. Целью гибридизации является попытка объединить преимущества экспертных правил и модели машинного обучения, повысить точность идентификации инцидентов, уменьшить количество ложных срабатываний и обеспечить динамическое обучение системы в условиях постоянно меняющейся угрозы информационной безопасности. Правила обеспечивают быстрый и прозрачный механизм реагирования на типовые сценарии, а также позволяют обеспечить понимание и объяснимость решений. Модели машинного обучения, в свою очередь, позволяют обнаруживать сложные, скрытые или неизвестные угрозы, новые классы атак и аномалий.

Структурная схема системы представлена на рисунке.

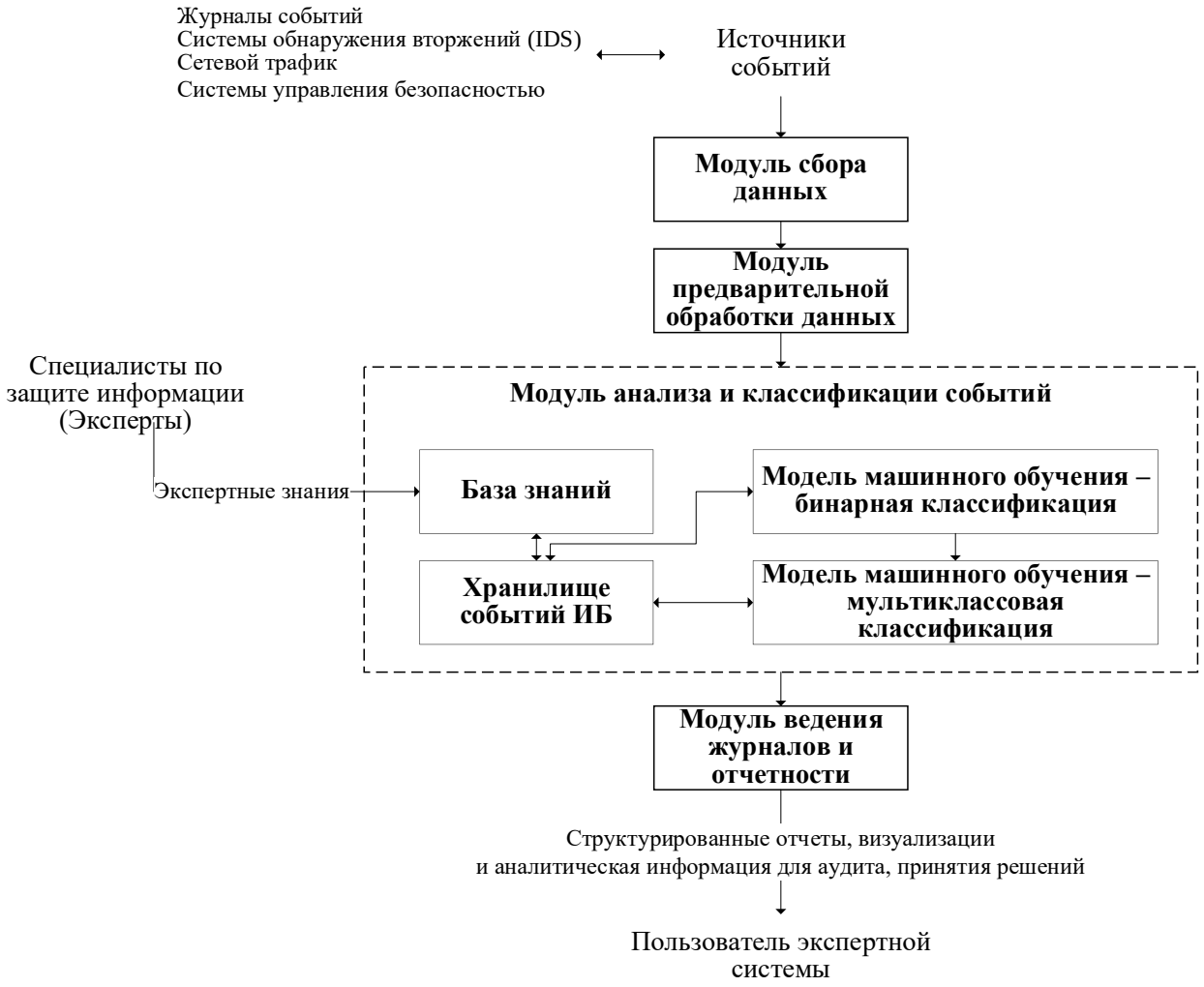

Рис. Структурная схема экспертной системы
Fig. Structural diagram of an expert system

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Предлагаемая архитектура экспертной системы представлена как концептуальная модель, включающая модули для сбора данных, предварительной обработки, анализа и классификации, а также ведения журналов, с использованием технологий реального времени, таких как Apache Kafka и Flink, для обеспечения масштабируемости и эффективности в производственной среде [14, 15].

Однако, для проведения экспериментальных исследований и валидации подхода был разработан прототип на языке программирования Python, позволяющий сосредоточиться на ключевых алгоритмах классификации без необходимости развертывания полной распределенной инфраструктуры. Выбор языка Python обусловлен тем фактом, что в его структуре имеется необходимый набор библиотек для машинного обучения и обработки данных, а также обеспечивается гибкость при интеграции с другими системами. Прототип использует локальные вычислительные ресурсы и библиотеки, такие как Pandas для манипуляции данными и Scikit-learn для реализации алгоритма Random Forest, что упрощает итеративное тестирование на доступных датасетах [16, 17].

Несмотря на различия в реализации, прототип инкапсулирует логику предлагаемой структуры системы, включая гибридизацию экспертных правил и классификации на основе машинного обучения, обеспечивая прямую связь между концепцией и эмпирическими результатами. Такой подход позволяет сосредоточиться на доказательстве эффективности методов на контролируемых

данных, а затем перейти к интеграции в распределенную систему, минимизируя затраты на ранних этапах разработки.

В качестве исходных данных для обучения и тестирования проектируемой экспертной системы было принято решение использовать набор данных UNSW-NB15, представляющий собой информационный ресурс в области исследования сетевой безопасности [18]. Набор данных предназначен для обнаружения вредоносного сетевого трафика с использованием методов машинного обучения.

В статье [19] набор данных UNSW-NB15 рассматривается в качестве автономного набора данных для разработки модели обнаружения вторжений с целью обнаружения вредоносных действий в сети. Авторы утверждают, что полученные оценки эффективности предлагаемой работы с UNSW-NB15 демонстрируют более высокую точность по сравнению с другими существующими подходами.

Общее количество записей (событий) в наборе UNSW-NB15 составляет 2540044, которые хранятся в четырех CSV-файлах. В датасете имеются сконфигурированные тренировочный (обучающий) и тестовый наборы данных, хранящиеся в файлах UNSW_NB15_training-set.csv и UNSW_NB15_testing-set.csv. Количество записей в тренировочном наборе составляет 175 341, в тестовом – 82 332 записи. Записи представляют собой состояния сетевого трафика, и включает разнообразные типы сетевых трафиков, охватывающие нормальную активность и сценарии девяти классов сетевых атак – Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode и Worms. Каждая запись характеризуется списком из 49 признаков, описывающих характеристики трафика (включая, метки классов, указывающие на класс атаки или отсутствие атаки для каждого события).

Методика обучения и тестирования экспертной системы на языке Python для анализа событий информационной безопасности состоит из следующих основных этапов:

1. Импорт библиотек (для чтения файлов CSV и манипуляции DataFrame, работы с массивами, масштабирования, кодировки меток, реализации функций оценки, организации модели деревьев решений).

2. Загрузка данных из CSV файлов (тренировочный и тестовый наборы данных).

3. Выбор ключевых признаков для анализа. Из 49 признаков событий, которые представлены в наборе данных UNSW-NB15, для демонстрации работы экспертной системы выбираются 10 признаков:

selected_features = ['proto', 'service', 'dur', 'sload', 'dload', 'sttl', 'dttl', 'sbytes', 'label', 'attack_cat']

При этом в данном случае признаки 'label', 'attack_cat' рассматриваются как целевые переменные классификации (метки для обучения и тестирования).

Выбор этих признаков обусловлен тем, что они обладают интерпретируемостью и релевантностью с точки зрения выявления сетевых атак из потока событий:

– `proto` (протокол): указывает на тип протокола (TCP, UDP, ICMP), что напрямую влияет на характер трафика;

– `service`: служебное приложение или сервис (HTTP, FTP, SSH), что помогает определить тип взаимодействия;

– dur` (длительность): время сессии — аномальные значения могут указывать на атаки;

– `sload`, `dload`: объем входящих/исходящих данных, показатель подозрительной активности;

– `sttl`, `dttl`: время существования пакета от источника к назначению/ от назначения к источнику, помогают идентифицировать фальсифицированный трафик или специфические атаки;

– `sbytes`: количество отправленных байт — индикатор активности.

Ограничение количества используемых для анализа признаков вызвано необходимостью реализации быстрого анализа с целью получения решения в реальном времени, а высокая размерность (49 признаков) замедляет обучение и принятия решения системой и требует больше вычислительных ресурсов. В итоге, выбор этих 10 признаков — это компромисс между полнотой данных и практичностью: они покрывают ключевые аспекты атак, позволяют строить эффективные правила.

4. Обработка категориальных и масштабирование числовых признаков. Категориальные признаки ('proto' и 'service') преобразуются в числовые с помощью one-hot encoding. Числовые признаки масштабируются в диапазон [0, 1] для нормализации.

5. Формирование признаков (X) и целевых переменных (y):

```
# Исключаем из признаков колонки 'label' и 'attack_cat'
X_train = training_set_encoded.drop(['label', 'attack_cat'], axis=1)
X_test = testing_set_encoded.drop(['label', 'attack_cat'], axis=1)

# Бинарная цель – 'label' (0 - Normal, 1 - Attack)
y_train_binary = training_set_encoded['label']
y_test_binary = testing_set_encoded['label']

# Мультиклассовая цель – категория атаки 'attack_cat'
label_encoder_y = LabelEncoder()
y_train_multiclass = label_encoder_y.fit_transform(training_set_encoded['attack_cat'])
y_test_multiclass = label_encoder_y.transform(testing_set_encoded['attack_cat'])
```

6. Создание и обучение моделей случайного леса (Random Forest) для бинарной (на тренировочном наборе данных) и мультиклассовой (из тренировочного набора данных исключаются нормальные события, определенные на этапе бинарной классификации) классификаций:

```
binary_model = RandomForestClassifier(n_estimators=200, random_state=42)
binary_model.fit(X_train, y_train_binary)

# Фильтруем тренировочные данные: только атаки (label == 1)
attack_train_indices = y_train_binary == 1
X_train_attacks = X_train[attack_train_indices]
y_train_multiclass_attacks = y_train_multiclass[attack_train_indices]

multiclass_model = RandomForestClassifier(n_estimators=150, random_state=42)
multiclass_model.fit(X_train_attacks, y_train_multiclass_attacks)
```

7. Реализация тестирования обученной модели Random Forest для бинарной классификации (на тестовом наборе данных):

```
# Предсказание бинарной модели: возвращает 0 для Normal, 1 для Attack
y_pred_binary = binary_model.predict(X_test)
# Получение вероятностей предсказаний: 2D-массив с вероятностями для Normal и Attack
probabilities_binary = binary_model.predict_proba(X_test)
# Определение индексов образцов, предсказанных как атаки (True для Attack)
attack_indices = y_pred_binary == 1
```

8. Реализация тестирования обученной модели Random Forest для мультиклассовой классификации (на сокращенном тестовом наборе данных):

```
predictions = ['Normal'] * n_samples
if np.any(attack_indices):
    X_test_attacks = X_test[attack_indices]
    y_pred_multiclass_attacks = multiclass_model.predict(X_test_attacks)
    probabilities_multiclass = multiclass_model.predict_proba(X_test_attacks)
    predictions_attack = [label_encoder_y.classes_[pred] for pred in y_pred_multiclass_attacks]
    for idx, attack_idx in enumerate(np.where(attack_indices)[0]):
        predictions[attack_idx] = predictions_attack[idx]
```

9. Реализация постобработки с помощью экспертных правил.

Предварительно был выполнен анализ результатов оценки эффективности системы с использованием только бинарной и мультиклассовой классификаций, с целью выявления классов атак, которые необходимо дополнительно уточнять:

Бинарная (Normal vs Attack) + Мультиклассовая (без применения правил)
Точность бинарной классификации: 0.890
Время предсказания (без правил): 3.13 сек
Точность: 0.781

Classification Report:

	precision	recall	f1-score	support
Analysis	0.00	0.00	0.00	677
Backdoor	0.04	0.06	0.05	583
DoS	0.42	0.14	0.21	4089
Exploits	0.57	0.85	0.68	11132
Fuzzers	0.32	0.47	0.38	6062
Generic	1.00	0.97	0.98	18871
Normal	0.93	0.82	0.87	37000
Reconnaissance	0.93	0.80	0.86	3496
Shellcode	0.27	0.39	0.32	378
Worms	0.50	0.48	0.49	44
accuracy			0.78	82332
macro avg	0.50	0.50	0.48	82332
weighted avg	0.81	0.78	0.79	82332

Confusion Matrix:

```
[[    0    46    37   593     0     0     1     0     0     0]
 [    0    37    23   510     3     0     8     0     2     0]
 [   33   372   554  2862   131    13    82    23    18     1]
 [   69   379   319  9431   385    13   293   178    56     9]
 [    0    62    86  1292  2841     5  1659     7   106     4]
 [    0     8    43   368    95 18294    52     2     7     2]
 [  144     2   213   957  5263    21 30201    12   184     3]
 [    6    66    26   513    42     2    35  2784    20     2]
 [    0     1     5    30   104     1    89     2   146     0]
 [    0     1     1    17     3     1     0     0     0    21]]
```

Модель Random Forest (бинарая + мультиклассовая, без применения правил) показывает общую точность 0,781 и время предсказания 3,13 сек. Внимание заслуживают классы с низким значением показателя recall, поскольку это указывает на проблемы с пропуском атак (false negatives), что критично для безопасности. Перечень трех атак, для которых наиболее целесообразно добавить правила постобработки (каскадные правила на оценках признаков dload, dur, sbytes, sload, sttl, dttl):

– Analysis (recall = 0.00): полное отсутствие предсказаний этого класса в тесте (все 677 образцов классифицированы неправильно, в основном как Exploits). Это редкий класс с пересечениями признаков, но его нулевой показатель recall делает его кандидатом на правила для улучшения обнаружения, хотя он менее критичный по сравнению с другими;

– Backdoor (recall = 0,06): очень низкий recall, с 583 образцами только 37 истинных положительных. Модель часто путает с Exploits и DoS. Правила могут повысить показатель recall, так как backdoor-атаки опасны для скрытого доступа;

– DoS (recall = 0,14): низкий показатель recall с 4089 образцами, но только 554 истинных положительных. Модель путает с Exploits и Fuzzers. Это критичный класс для сетевой безопасности, и правила помогут фокусироваться на паттернах, таких как высокий признак dload и низкий признак dur.

Постобработка применяется только для предсказаний этих классов с вероятностью ниже порога (например, <0.8), чтобы корректировать неуверенные решения модели. Это делает подход целенаправленным и не влияет на уверенные предсказания других классов.

Для построения правил используются переменные экспертной системы, в качестве которых выступают признаки событий. Правила базы знаний экспертной системы – это ключевые компоненты, которые формируют логику и знания, используемые системой для принятия решений и вывода рекомендаций. Правила представляют собой формализованные логические конструкции, которые связывают условия (факты или признаки) с выводами или действиями, и записываются в виде «если – то» (если условие выполнено, то делается вывод или предпринимается действие).

Экспертные правила, используемые для постобработки предсказаний модели Random Forest, были разработаны на основе эмпирического анализа статистических характеристик тренировочного и тестового наборов данных UNSW-NB15. Этот анализ включал вычисление квартилей (25-й и 75-й процентилей), средних значений, модальных распределений, а также оценку пересечений между классами для минимизации ложных срабатываний. Процесс был итеративным: сначала проводился разведочный анализ данных (EDA) для выявления дискриминационных признаков (dload, dur, sbytes, sload, sttl, dttl), затем определялись пороги на основе квартилей, чтобы правила отражали типичные паттерны атак, одновременно учитывая дисбаланс классов.

Программный код реализации экспертных правила:

analysis_condition = (sample['dttl'] > 0.9) & (sample['dload'] < 0.0005) & (sample['sbytes'] < 0.0001)

dos_condition = (sample['dttl'] < 0.01) & (((sample['dload'] > 0.005) | (sample['sbytes'] > 0.05) | (sample['sload'] > 0.05)) | ((sample['dur'] < 1.0e-7) | (sample['dur'] > 0.5)))

backdoor_condition = (sample['dttl'] < 0.01) & (sample['dload'] < 0.0001) & (sample['sbytes'] < 0.00003) & (sample['sload'] < 0.00001)

Таким образом, предсказания и вероятности для каждого образца, полученные после выполнения предварительной классификации с помощью мультиклассовой модели, служат базой для дальнейшей корректировки. Особое внимание уделяется классам с низким показателем recall – DoS, Backdoor и Analysis, поскольку модель в этих случаях ошибается чаще всего.

10. Подсчет скорректированных предсказаний и оценка точности экспертной системы. Подсчитывается количество образцов, скорректированных постобработкой правилами для каждого класса атак (Analysis, Backdoor, DoS). Затем предсказания преобразуются в числовой формат с помощью label_encoder, измеряется время выполнения предсказания и выводится итоговая точность модели. В качестве результатов выводится подробный отчет классификации (precision, recall, f1-score) и матрица ошибок для оценки качества предсказаний после объединения.

Оценка точности для предложенной гибридной системы, сочетающей бинарную, мультиклассовую классификации и постобработку с помощью экспертных правил:

Бинарная + Мультиклассовая классификация + постобработка правилами (поиск угроз среди других классов, с порогом уверенности)
Шаг 1: Бинарная + Мультикласс предсказали все 82332 образцов
Шаг 2: Постобработка правилами скорректировала (с порогом 0.8): 1957 Analysis, 2777 DoS, 146 Backdoor

Время предсказания (с правилами): 212.34 сек
Точность: 0.769

Classification Report (после объединения):

	precision	recall	f1-score	support
Analysis	0.01	0.02	0.01	677
Backdoor	0.04	0.09	0.06	583
DoS	0.33	0.33	0.33	4089
Exploits	0.64	0.71	0.67	11132
Fuzzers	0.32	0.47	0.38	6062
Generic	1.00	0.97	0.98	18871
Normal	0.93	0.82	0.87	37000

Reconnaissance	0.98	0.74	0.84	3496
Shellcode	0.33	0.26	0.29	378
Worms	0.64	0.16	0.25	44
accuracy			0.77	82332
macro avg	0.52	0.46	0.47	82332
weighted avg	0.82	0.77	0.79	82332

Confusion Matrix (после объединения):

```
[[   13    53   260   350     0     0     1     0     0     0]
 [    7    50   232   283     3     0     8     0     0     0]
 [  107   436  1330  1969   131    13    82     9    12     0]
 [  849   420  1211  7901   385    13   293    41    17     2]
 [   82    73   560   770  2841     5  1659     2    70     0]
 [   19    15    58   331    95 18294    52     1     4     2]
 [  659     2   292   464  5263    21 30201     4    94     0]
 [  398    69   112   237    42     2    35  2598     3     0]
 [   54     1    26     4   104     1    89     1    98     0]
 [   21     1     3     8     3     1     0     0     0     7]]
```

Общая точность предсказаний снижается с 0,781 (классификация без правил) до 0,769 (классификация с правилами) на 1,2%, что обусловлено переклассификацией 1957 Analysis, 2777 DoS и 146 Backdoor, приводящей к ложным срабатываниям в доминирующих классах, но это приемлемо для приоритета полноты в обнаружении угроз (минимизации пропущенных угроз).

Для класса Analysis показатель recall растет с 0% до 2% (13 истинных положительных), показетль precision остается 0,01, улучшая обнаружение за счет переклассфикации из Exploits.

Класс Backdoor показывает рост показателя recall с 6% (37 истинных) до 9% (50 истинных), показатель precision стабилен на 0,04.

Класс DoS демонстрирует значительный рост показателя recall с 14% (554 истинных) до 33% (1330 истинных), но показатель precision падает с 0,42 до 0,33 из-за ложных положительных.

Время выполнения растет с 3,13 сек до 212,34 сек (в ~68 раз), требуя оптимизации постобработки.

Предложенная гибридная архитектура экспертной системы, основанная на сочетании машинного обучения и экспертных правил, демонстрирует свою эффективность в сценариях с дисбалансом классов атак, где чистое машинное обучение пропускает редкие атаки (показатель recall <15% для DoS, 0% для Analysis). Правила компенсируют это, используя знания о специфических значениях признаков атак, что повышает полноту на 19–100% для целевых классов без полного переобучения модели.

Преимуществом такого решения также является интерпретируемость полученных предсказаний, поскольку правила объяснимы. В качестве недостатков следует отметить снижение точности, и низкая производительность из-за итеративного применения правил

ЗАКЛЮЧЕНИЕ

Разработанная гибридная структура экспертной системы на основе машинного обучения и набора экспертных правил подтвердила свою работоспособность в задаче выявления угроз информационной безопасности, обеспечивая интерпретируемый анализ сетевых событий. Объединение машинного обучения и экспертных правил, основанных на выбранных признаках (proto, service, dur, sload, dload, sttl, dttl), позволило создать прозрачную модель, которая эффективно классифицирует типы атак и нормальных событий, а также обеспечивает возможность объяснения принятых решений. Такой подход способствует контролю и пониманию процесса обнаружения, что особенно важно в условиях ограниченных данных или необходимости быстрого реагирования.

Результаты экспериментов на прототипе, реализованном на языке программирования Python, оправдывают дальнейшую разработку полной системы. Предложенная гибридная модель экспертной системы демонстрирует прогресс в классификации сетевых атак на датасете UNSW-NB15, особенно для редких классов, где показатель recall улучшается на 19–100% при приемлемом снижении общей точности (на 1,2%). Этот подход балансирует автоматизацию процессов на основе машинного обучения с человеческой экспертизой, делая его подходящим для критических приложений безопасности. Результаты подтверждают, что гибридные методы представляют собой рациональный и целесообразный путь к повышению надежности систем обнаружения вторжений (IDS) в условиях дисбаланса данных.

В качестве направления дальнейшего исследования и развития экспертной системы предлагается сфокусироваться на оптимизации правил, например, через обучение с подкреплением (Reinforcement learning), и интеграции с другими моделями, такими как LSTM для анализа временных рядов, чтобы минимизировать компромиссы между точностью и производительностью [20]. Также необходимо решить задачу автоматического обновления и расширения правил на основе новых данных, реализовать автоматическую адаптацию системы к изменяющейся ситуации в сети. Планируется интеграция обучающих алгоритмов для повышения точности классификации, развитие интерфейса для повышения интерактивности и удобства использования, а также внедрение механизмов объяснения решений, что повысит доверие пользователей и облегчит интерпретацию результатов.

Список литературы

1. Seraphim B.I., Palit Sh., Srivastava K., Poovammal E. A Survey on Machine Learning Techniques in Network Intrusion Detection System. – 2018. – pp. 1-5. DOI: 10.1109/CCAA.2018.8777596.
2. Mua U.S., Chakraborty S., Abdullahi M.M., Maini T. A Review on Intrusion Detection System using Machine Learning Techniques, 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021. – pp. 541-549, DOI: 10.1109/ICCCIS51004.2021.9397121.
3. Prajapati A., Gupta Sh. A Survey: Data Mining and Machine Learning Methods for Cyber Security. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. – 2021. – pp. 24-34. DOI: 10.32628/CSEIT217212.
4. Veeramreddy J., Prasad K. Anomaly-Based Intrusion Detection System. – 2019. DOI: 10.5772/intechopen.82287.
5. Dutta A. Random Forest Classifier Based Network Intrusion Detection System. Engineering, Technology and Applied Science Research. – 2021. – No 9. – Pp. 4603-4608. DOI: 10.22214/ijraset.2021.35406.
6. Abdelaziz M.T., Radwan A., Mamdouh H. et al. Enhancing Network Threat Detection with Random Forest-Based NIDS and Permutation Feature Importance. J Netw Syst Manage. – 2025. – 33, 2. https://doi.org/10.1007/s10922-024-09874-0
7. Prajapati P.K., Singh I., Subhashini N. Network Intrusion Detection Using Machine Learning. In: Subhashini N., Ezra M.A.G., Liaw SK. (eds) Futuristic Communication and Network Technologies. Lecture Notes in Electrical Engineering, vol 966. Springer, Singapore. – 2023. https://doi.org/10.1007/978-981-19-8338-2_4
8. Sowmya T., Anita M. A novel stable feature selection algorithm for machine learning based intrusion detection system. Procedia Computer Science. – 2025. – P. 252. 738-747. DOI: 10.1016/j.procs.2025.01.034.
9. Бабичева М.В., Третьяков И.А. Применение методов машинного обучения для автоматизированного обнаружения сетевых вторжений. Вестник Дагестанского государственного технического университета. Технические науки. – 2023. – 50(1). – С. 53-61. https://doi.org/10.21822/2073-6185-2023-50-1-53-61
10. Гайдук К.А. К вопросу о реализации алгоритмов выявления внутренних угроз с применением машинного обучения / К.А. Гайдук, А.Ю. Исхаков // Вестник СибГУТИ. – 2022. – Т. 16, № 4. – С. 80-95. – DOI: 10.55648/1998-6920-2022-16-4-80-95. – EDN SGBSIH.
11. Перспективные направления применения технологий искусственного интеллекта при защите информации / Р.В. Мещеряков, С.Ю. Мельников, В.А. Пересыпкин, А.А. Хорев // Вопросы кибербезопасности. – 2024. – № 4(62). – С. 2-12. – DOI 10.21681/2311-3456-2024-4-02-12. – EDN GJWQWP.
12. Селеменев А.В. Применение искусственных иммунных систем для обнаружения сетевых вторжений / А.В. Селеменев, И.Ф. Астахова, Е.В. Трофименко // Вестник Воронежского государственного

университета. Серия: Системный анализ и информационные технологии. – 2019. – № 2. – С. 49-56. – EDN XDMKTQ.

13. Котенко И.В. Система сбора, хранения и обработки информации и событий безопасности на основе средств Elastic Stack / И.В. Котенко, А.А. Кулешов, И.А. Ушаков // Труды СПИИРАН. – 2017. – № 5(54). – С. 5-34. – DOI 10.15622/sp.54.1. – EDN ZMREVZ.

14. Carbone P., Katsifodimos A., Ewen S., Markl V., Haridi S., Tzoumas K. Apache Flink™: Stream and Batch Processing in a Single Engine. IEEE Data Engineering Bulletin. – 2015. – 38(4). – pp. 28-38 p.

15. Daksa R., Kemala A. A Comparative Study on Real Time Data Streaming for Fraud Detection Using Kafka with Apache Flink and Apache Spark. Procedia Computer Science. – 2025. – 269. – pp. 192-199. DOI: 10.1016/j.procs.2025.08.272.

16. Diana Julie M.D. Exploring the Paradigm Shift: Harnessing Data Analytics for Real – World Applications / D. Diana Julie M // International Journal of Science and Research. – 2023. – Vol. 12, No. 6. – P. 1467-1480. – DOI: 10.21275/sr23611121501. – EDN IVRRST.

17. Редченков Д.С. Библиотека Pandas для анализа данных в Python / Д.С. Редченков, Д.И. Ильин // Информационно-вычислительные технологии и их приложения: Сборник статей XXIX Международной научно-технической конференции, Пенза, 15–16 августа 2025 года. – Пенза: Пензенский государственный университет архитектуры и строительства, 2025. – С. 178-182. – EDN JQCIMM.

18. Moustafa N., Jill S. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) 2015 Military Communications and Information Systems Conference (MilCIS). Canberra, ACT, – 2015. pp. 1-6. DOI: 10.1109/MilCIS.2015.7348942

19. Kumar V., Das A., Sinha D. Statistical Analysis of the UNSW-NB15 Dataset for Intrusion Detection. – 2020. DOI: 10.1007/978-981-13-9042-5_24.

20. Meliboyev A. Long Short Term Memory Algorithm in Intrusion Detection: A Deep Learning Approach to Time Series Data. SSRN Electronic Journal. – 2025. DOI: 10.2139/ssrn.5527203.

References

1. Seraphim B.I., Palit Sh., Srivastava K., Poovammal E. A Survey on Machine Learning Techniques in Network Intrusion Detection System. – 2018. – pp. 1-5. DOI: 10.1109/CCAA.2018.8777596.

2. Mua U.S., Chakraborty S., Abdullahi M.M., Maini T. A Review on Intrusion Detection System using Machine Learning Techniques, 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021. – pp. 541-549, DOI: 10.1109/ICCCIS51004.2021.9397121.

3. Prajapati A., Gupta Sh. A Survey: Data Mining and Machine Learning Methods for Cyber Security. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. – 2021. – pp. 24-34. DOI: 10.32628/CSEIT217212.

4. Veeramreddy J., Prasad K. Anomaly-Based Intrusion Detection System. – 2019. DOI: 10.5772/intechopen.82287.

5. Dutta A. Random Forest Classifier Based Network Intrusion Detection System. Engineering, Technology and Applied Science Research. – 2021. – No 9. – Pp. 4603-4608. DOI: 10.22214/ijraset.2021.35406.

6. Abdelaziz M.T., Radwan A., Mamdouh H. et al. Enhancing Network Threat Detection with Random Forest-Based NIDS and Permutation Feature Importance. J Netw Syst Manage. – 2025. – 33, 2. https://doi.org/10.1007/s10922-024-09874-0

7. Prajapati P.K., Singh I., Subhashini N. Network Intrusion Detection Using Machine Learning. In: Subhashini N., Ezra M.A.G., Liaw SK. (eds) Futuristic Communication and Network Technologies. Lecture Notes in Electrical Engineering, vol 966. Springer, Singapore. – 2023. https://doi.org/10.1007/978-981-19-8338-2_4

8. Sowmya T., Anita M. A novel stable feature selection algorithm for machine learning based intrusion detection system. Procedia Computer Science. – 2025. – P. 252. 738-747. DOI: 10.1016/j.procs.2025.01.034.

9. Babicheva M.V., Tretyakov I.A. Application of machine learning methods for automated detection of network intrusions. Herald of Dagestan State Technical University. Technical Sciences. – 2023. – 50(1). – pp. 53-61. (In Russ.) https://doi.org/10.21822/2073-6185-2023-50-1-53-61

10. Gaiduk K.A., Iskhakov A.Yu. On the Implementation of Algorithms for Detecting Insider Threats Using Machine Learning. Vestnik SibGUTI. – 2022. – 16. – No. 4. – pp. 80-95. https://doi.org/10.55648/1998-6920-2022-16-4-80-95.

11. Meshcheryakov R.V., Melnikov S.Yu., Peresypkin V.A., Khorev A.A. Promising Directions for the Application of Artificial Intelligence Technologies in Information Security. Cybersecurity Issues. – 2024. – No. 4(62). – pp. 2-12. https://doi.org/10.21681/2311-3456-2024-4-02-12.

12. Selemenev A.V., Astakhova I.F., Trofimenko E.V. Application of Artificial Immune Systems for Detecting Network Intrusions. Vestnik of Voronezh State University. Series: System Analysis and Information Technologies. – 2019. – No. 2. – pp. 49–56.

13. Kotenko I.V., Kuleshov A.A., Ushakov I.A. System for Collecting, Storing and Processing Security Information and Events Based on Elastic Stack Tools. Proceedings of SPIIRAN. – 2017. – No. 5(54). – pp. 5-34. https://doi.org/10.15622/sp.54.1.

14. Carbone P., Katsifodimos A., Ewen S., Markl V., Haridi S., Tzoumas K. Apache Flink™: Stream and Batch Processing in a Single Engine. IEEE Data Engineering Bulletin. – 2015. – 38(4). – pp. 28-38 p.

15. Daksa R., Kemala A. A Comparative Study on Real Time Data Streaming for Fraud Detection Using Kafka with Apache Flink and Apache Spark. Procedia Computer Science. – 2025. – 269. – pp. 192-199. DOI: 10.1016/j.procs.2025.08.272.

16. Diana Julie M.D. Exploring the Paradigm Shift: Harnessing Data Analytics for Real – World Applications / D. Diana Julie M // International Journal of Science and Research. – 2023. – Vol. 12, No. 6. – P. 1467-1480. – DOI: 10.21275/sr23611121501. – EDN IVRRST.

17. Redchenkov D.S., Ilin D.I. Pandas Library for Data Analysis in Python. In Information and Computational Technologies and Their Applications: Collection of Articles of the XXIX International Scientific and Technical Conference, Penza, August 15–16, 2025. – pp. 178–182. Penza: Penza State University of Architecture and Construction, 2025.

18. Moustafa N., Jill S. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) 2015 Military Communications and Information Systems Conference (MilCIS). Canberra, ACT, – 2015. pp. 1-6. DOI: 10.1109/MilCIS.2015.7348942

19. Kumar V., Das A., Sinha D. Statistical Analysis of the UNSW-NB15 Dataset for Intrusion Detection. – 2020. DOI: 10.1007/978-981-13-9042-5_24.

20. Meliboyev A. Long Short Term Memory Algorithm in Intrusion Detection: A Deep Learning Approach to Time Series Data. SSRN Electronic Journal. – 2025. DOI: 10.2139/ssrn.5527203.

Потиенко Даниил Анатольевич, магистрант кафедры «Информационная безопасность в вычислительных системах и сетях», ФГБОУ ВО «Донской государственный технический университет» (ДГТУ), г. Ростов-на-Дону, Россия
Газизов Андрей Равильевич, кандидат педагогических наук, доцент, доцент кафедры «Информационная безопасность в вычислительных системах и сетях», ФГБОУ ВО «Донской государственный технический университет» (ДГТУ), г. Ростов-на-Дону, Россия
Легонько Ольга Леонидовна, кандидат технических наук, доцент, доцент кафедры «Информационная безопасность в вычислительных системах и сетях», ФГБОУ ВО «Донской государственный технический университет» (ДГТУ), г. Ростов-на-Дону, Россия

Potienko Daniil Anatolyevich, Master's student of the Department of Information security in computing systems and networks, Don State Technical University, Rostov-on-Don, Russia
Gazizov Andrey Ravilevich, Candidate of Pedagogical Sciences, Associate Professor, Associate Professor of the Department of Information security in computing systems and networks, Don State Technical University, Rostov-on-Don, Russia
Legonko Olga Leonidovna, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Information security in computing systems and networks, Don State Technical University, Rostov-on-Don, Russia