MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН

KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ — INTERNATIONAL UNIVERSITY — SINCE 2009 — МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Published since 2020.
Volume 7. 1 (25). 2026
January–March

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

2020 жылдан бері шығарылады
Том 7. 1 (25). 2026
Қаңтар-Наурыз

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

Издается с 2020 г.
Том 7. 1 (25). 2026
Январь-Март

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № KZ82VPY00020475, выданное от 20.02.2020 г.

Зарегистрировано в Международном центре регистрации серийных изданий ISSN (ЮНЕСКО, Париж, Франция). ISSN 2708–2032 (print), ISSN 2708–2040 (online)

Журнал входит в Перечень научных изданий, рекомендуемых КОКНВО МНВО РК для публикации основных результатов научной деятельности.

International Journal of Information and Communication Technologies
Periodicity: 4 times a year.
Languages: Kazakh, Russian, English
DOI prefix: 10.54309
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Thematic focus: "Information technology"; "Digital technologies in the development of socio-economic systems"; "Information security and communication technologies".

Journal website: https://journal.iitu.edu.kz
Owner: International Information Technology University JSC (Almaty).

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Префикс DOI: 10.54309
Мерзімділігі: жылына 4 рет.
Басылым тілі: қазақ, орыс, ағылшын.
Тақырып бағыты: "Ақпараттық технологиялар"; "Ақпараттық қауіпсіздік және коммуникациялық технологиялар"; "Әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология".
Журнал сайты: https://journal.iitu.edu.kz

Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Префикс DOI: 10.54309
Периодичность: 4 выпусков в год.
Язык издания: казахский, русский, английский.
Тематическая направленность: "Информационные технологии"; "Информационная безопасность и коммуникационные технологии"; "Цифровые технологии в развитии социо-экономических систем".
Сайт журнала: https://journal.iitu.edu.kz

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

CONTENTS

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

INFORMATION TECHNOLOGY

INFORMATION SECURITY AND COMMUNICATION TECHNOLOGIES

МАЗМҰНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМЫТУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

СОДЕРЖАНИЕ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИСОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.1. Number 25 (2026). Pp. 292–310
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.25.1.018
УДК 004.056.5

DETECTION OF CYBER ATTACKS IN TRANSPORT NETWORKS BASED ON MACHINE LEARNING METHODS

N.E. Karabayev[1], S.K. Serikbayeva[1]*, Y.M. Mardenov[2], B. Tassuov[3], M. Fajkus[4]
[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[2] Astana International University, Astana, Kazakhstan;
[3] Taraz University named after M.Kh. Dulaty, Taraz, Kazakhstan;
[4] Tomas Bata University in Zlín, Zlín, Czech Republic.
E-mail: inf_8585@mail.ru

Nurdaulet E. Karabayev — Doctoral student, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0009-0008-6532-6382;
Sandugash K. Serikbayeva — PhD, Senior Lecturer, Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
E-mail: inf_8585@mail.ru, https://orcid.org/0000-0002-3627-3321;
Yerik M. Mardenov — Ms.Sc., Higher School of Information Technologies and Engineering, Astana International University, Astana, Kazakhstan
https://orcid.org/0000-0001-9284-9797**;**
Bolat Tassuov — Associate Professor, Department of Physics and Informatics, Taraz University named after M.Kh. Dulaty, Taraz, Kazakhstan
https://orcid.org/0000-0002-2000-6720;
Martin Fajkus — PhD, Senior Lecturer, Faculty of Applied Informatics, Tomas Bata University in Zlín, Zlín, Czech Republic
https://orcid.org/0000-0002-5698-1106.

Abstract. With the digitalization of vehicles and the growth of the number of electronic control units, ensuring the cybersecurity of automotive networks is becoming one of the priority tasks. Modern vehicles are complex cyberphysical systems in which data exchange between electronic components is carried out via the CAN (Controller Area Network) bus. Despite the widespread adoption and reliability of this protocol, the CAN architecture did not initially provide mechanisms

for protection against cyber attacks, which makes transport networks vulnerable to various types of intervention, including attacks such as DoS, Fuzzy, RPM Spoofing and Gear Spoofing. This paper discusses the task of automatically detecting and classifying cyberattacks in automotive networks based on machine learning methods. The open car hacking dataset was used as the initial data, containing real logs of CAN messages both under normal conditions and when simulating attacks. Preliminary data processing was performed, including cleaning, normalization and balancing of classes, as well as analysis of feature correlation. To solve the multiclass classification problem, two machine learning algorithms were implemented and compared: XGBoost and logistic regression. The quality of the models was assessed using error matrices and accuracy analysis by class. The results of the experiments showed that the XGBoost model demonstrates higher accuracy and robustness in classifying attacks compared to logistic regression, especially for most attacking classes. Additional analysis of the importance of the features made it possible to identify the most informative parameters of CAN messages, reflecting the nature of the injected attacks. The results confirm the effectiveness of the application of machine learning methods to improve the level of security of transport networks and can be used in the development of intelligent intrusion detection systems in car CAN networks.

Keywords: cybersecurity, transport networks, CAN-bus, cyberattacks, machine learning, XGBoost, logistic regression, anomaly detection, attack classification, automotive networks

For citation: N. Karabayev, S. Serikbayeva, Y. Mardenov, B. Tassuov, M. Fajkus (2026). Detection of cyber attacks in transport networks based on machine learning methods // International journal of information and communication technologies. Vol. 7. No.25. Pp. 292-310. https://doi.org/10.54309/ijict.2026.25.1.018. (In Russ.).

Conflict of interest: The authors declare that there is no conflict of interest.

МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН КӨЛІК ЖЕЛІЛЕРІНДЕГІ КИБЕРШАБУЫЛДАРДЫ АНЫҚТАУ

Н.Е. Қарабаев[1], С.К. Серикбаева[1]*, Е.М. Марденов[2], Б. Тасуов[3], М. Файкус[4]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

[2]Астана халықаралық университеті, Астана, Қазақстан;

[3]М. Х. Дулати атындағы Тараз университеті , Тараз, Қазақстан;

[4]Злиндегі Томас Бата университеті, Злин , Чех Республикасы.

E-mail: inf_8585@mail.ru

Қарабаев Нұрдәулет Ерланұлы — докторант, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан

https://orcid.org/0009-0008-6532-6382;

Серикбаева Сандугаш Курманбековна — PhD, ақпараттық жүйелер

кафедрасының аға оқытушысы, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
E-mail: inf_8585@mail.ru, https://orcid.org/0000-0002-3627-3321;
Марденов Ерік Маратұлы — магистр, Ақпараттық технологиялар және инженерия жоғары мектебі, Астана халықаралық университеті, Астана, Қазақстан https://orcid.org/0000-0001-9284-9797;
Тасуов Болат — физика және информатика кафедрасының қауысдастырылған профессоры, М. Х. Дулати атындағы Тараз университеті , Тараз, Қазақстан https://orcid.org/0000-0002-2000-6720;
Файкус Мартин — PhD, қолданбалы информатика факультетінің аға оқытушысы, Злиндегі Томас Бата университеті, Злин , Чех Республикасы https://orcid.org/0000-0002-5698-1106.

Аннотация. Көлік құралдарын цифрландыру және басқарудың электрондық блоктары санының өсуі жағдайында автомобиль желілерінің киберқауіпсіздігін қамтамасыз ету басым міндеттердің біріне айналуда. Қазіргі заманғы көлік құралдары электрондық компоненттер арасында деректер алмасу CAN (Controller Area Network) шинасы бойынша жүзеге асырылатын күрделі киберфизикалық жүйелер болып табылады. Осы хаттаманың кең таралуына және сенімділігіне қарамастан, CAN архитектурасы бастапқыда кибершабуылдардан қорғау тетіктерін көздемеген, бұл көлік желілерін DoS, Fuzzy, RPM Spoofing және Gear Spoofing сияқты шабуылдарды қоса алғанда, араласудың әртүрлі түрлеріне осал етеді. Бұл жұмыста машиналық оқыту әдістері негізінде автомобиль желілерінде киберқақтарды автоматты түрде анықтау және жіктеу міндеті қарастырылады. Бастапқы деректер ретінде қалыпты жағдайларда да, шабуылдарды модельдеу кезінде де CAN-хабарламалардың нақты журналдарын қамтитын Car Hacking Dataset ашық жинағы пайдаланылды. Сыныптарды тазартуды, қалыпқа келтіруді және теңгерімдеуді, сондай-ақ белгілердің корреляциясын талдауды қамтитын деректерді алдын ала өңдеу жүргізілді. Мультиклассалық жіктеу міндетін шешу үшін машиналық оқытудың екі алгоритмі іске асырылды және салыстырылды: XGBoost және логистикалық регрессия. Модельдер сапасын бағалау қателер матрицаларын және сыныптар бойынша дәлдікті талдауды пайдалана отырып жүргізілді. Эксперименттердің нәтижелері көрсеткендей, XGBoost моделі логистикалық регрессиямен салыстырғанда, әсіресе көптеген шабуылдаушы кластар үшін шабуылдарды жіктеу кезінде жоғары дәлдік пен тұрақтылықты көрсетеді. Белгілердің маңыздылығын қосымша талдау инжектирленетін шабуылдардың сипатын көрсететін CAN-хабарламалардың неғұрлым ақпараттық параметрлерін анықтауға мүмкіндік берді. Алынған нәтижелер көлік желілерінің қауіпсіздік деңгейін арттыру үшін машиналық оқыту әдістерін қолдану тиімділігін растайды және CAN-желілерінде басып

кіруді анықтаудың зияткерлік жүйелерін әзірлеу кезінде пайдаланылуы мүмкін.

Түйін сөздер: киберқауіпсіздік, көлік желілері, CAN-шина, кибер шабуылдар, машиналық оқыту, XGBoost, логистикалық регрессия, ауытқуларды анықтау, шабуылдарды жіктеу, автомобиль желілері

Дәйексөздер үшін: Н.Е. Қарабаев, С.К. Серикбаева, Е.М. Марденов, Б. Тасуов, М. Файкус (2026). Машиналық оқыту әдістеріне негізделген көлік желілеріндегі кибершабуылдарды анықтау // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т 7. № 25. Б. 292-310. https://doi.org/10.54309/IJICT.2026.25.1.018. (Орыс тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ОБНАРУЖЕНИЕ КИБЕРАТАК В ТРАНСПОРТНЫХ СЕТЯХ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Н.Е. Карабаев[1], С.К. Серикбаева[1*], Е.М. Марденов[2], Б. Тасуов[3], М. Файкус[4]

[1]Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан;
[2]Международный университет Астана, Астана, Казахстан;
[3]Таразский университет им. М.Х. Дулати, Тараз, Казахстан;
[4]Университет Томаса Баты в Злине, Злин, Чешская Республика.
E-mail: inf_8585@mail.ru

Карабаев Нурдаулет Ерланович — докторант, Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан
https://orcid.org/0009-0008-6532-6382;
Серикбаева Сандугаш Курманбековна — PhD, старший преподаватель кафедры информационных систем, Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан
E-mail: inf_8585@mail.ru, https://orcid.org/0000-0002-3627-3321;
Марденов Ерик Маратович — магистр, Высшая школа информационных технологий и инженерии, Международный университет Астана, Астана, Казахстан
https://orcid.org/0000-0001-9284-9797;
Тасуов Болат — ассоциированный профессор, кафедра физики и информатики, Таразский университет имени М.Х. Дулати, Тараз, Казахстан
https://orcid.org/0000-0002-2000-6720;
Файкус Мартин — PhD, старший преподаватель, факультет прикладной информатики, Университет Томаша Баты в Злине, Злин, Чешская Республикаhttps://orcid.org/0000-0002-5698-1106.

Аннотация. В условиях цифровизации транспортных средств и роста

числа электронных блоков управления обеспечение кибербезопасности автомобильных сетей становится одной из приоритетных задач. Современные транспортные средства представляют собой сложные киберфизические системы, в которых обмен данными между электронными компонентами осуществляется по шине CAN (Controller Area Network). Несмотря на широкое распространение и надежность данного протокола, архитектура CAN изначально не предусматривала механизмов защиты от кибератак, что делает транспортные сети уязвимыми к различным видам вмешательства, включая атаки типа DoS, Fuzzy, RPM Spoofing и Gear Spoofing. В данной работе рассматривается задача автоматического обнаружения и классификации кибератак в автомобильных сетях на основе методов машинного обучения. В качестве исходных данных использовался открытый набор Car Hacking Dataset, содержащий реальные журналы CAN-сообщений как в нормальных условиях, так и при моделировании атак. Проведена предварительная обработка данных, включающая очистку, нормализацию и балансировку классов, а также анализ корреляции признаков. Для решения задачи мультиклассовой классификации были реализованы и сравнены два алгоритма машинного обучения: XGBoost и логистическая регрессия. Оценка качества моделей проводилась с использованием матриц ошибок и анализа точности по классам. Результаты экспериментов показали, что модель XGBoost демонстрирует более высокую точность и устойчивость при классификации атак по сравнению с логистической регрессией, особенно для большинства атакующих классов. Дополнительный анализ важности признаков позволил выявить наиболее информативные параметры CAN-сообщений, отражающие характер инжектируемых атак. Полученные результаты подтверждают эффективность применения методов машинного обучения для повышения уровня безопасности транспортных сетей и могут быть использованы при разработке интеллектуальных систем обнаружения вторжений в автомобильных CAN-сетях.

Ключевые слова: кибербезопасность, транспортные сети, CAN-шина, кибератаки, машинное обучение, XGBoost, логистическая регрессия, обнаружение аномалий, классификация атак, автомобильные сети

Для цитирования: Н.Е. Карабаев, С.К. Серикбаева, Е.М. Марденов, Б. Тасуов, М. Файкус (2026). Обнаружение кибератак в транспортных сетях на основе методов машинного обучения // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 25. Стр. 292–310. https://doi.org/10.54309/IJICT.2026.25.2.018.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Финансирование. *Настоящая работа сотрудниками НАО «Евразийский национальный университет имени Л.Н. Гумилева» проводится при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан (Грант № AP23489127).*

Введение.

Современные транспортные средства представляют собой сложные киберфизические системы с множеством электронных блоков управления (ECU). ECU отвечают за взаимодействие различных подсистем автомобиля, таких как двигатель, трансмиссия, тормозная система, система стабилизации и мультимедийные и вспомогательные услуги. Шина контроллерной сети (CAN) используется для эффективного обмена информацией между этими частями. Благодаря своей универсальности, надежности и низкой стоимости внедрения она на сегодняшний день стала стандартом в индустрии. Тем не менее, при разработке архитектуры CAN основное внимание уделяли производительности и устойчивости передачи данных в условиях ограниченных вычислительных ресурсов, а проблемы информационной безопасности оставались второстепенными. Следовательно, CAN-сети обладают значительными уязвимостями, которые активно используются в современных сценариях кибератак несмотря на то, что они широко распространены и проверены временем. С ростом количества электронных систем в транспортных средствах и возможностью удаленного взаимодействия с автомобилем посредством беспроводных интерфейсов исследования безопасности CAN-сетей становятся все более важными. Это связано с увеличением поверхности атак и рисков для пользователей.

Кибератаки на автомобильные сети могут сильно повлиять на функционирование автомобилей. Атаки типа отказа в обслуживании (DoS) (Cil, 2021), которые блокируют передачу данных между электронными блоками за счет массовой инъекции сообщений, являются одними из наиболее распространенных угроз. Другие типы угроз включают атаки тумана, которые дестабилизируют работу сети, передавая случайные идентификаторы и данные; и атаки на подмену, которые подменяют важные параметры, такие как обороты двигателя (RPM) или положение передачи (Gear) (Xiao et all., 2019).

Такие атаки могут привести к ложному отображению информации на приборной панели или полной потере управляемости автомобилем, поэтому особенно важно своевременно обнаруживать и классифицировать их. Из-за ограниченных вычислительных ресурсов ECU, требований к минимальной задержке передачи данных и необходимости поддерживать совместимость с существующими протоколами традиционные методы защиты, такие как криптографические механизмы или системы контроля доступа, неэффективны в условиях автомобильных сетей. В результате исследователи все больше обращают внимание на методы анализа данных и машинного обучения, которые позволяют выявлять закономерности в поведении сети и автоматически обнаруживать аномалии, связанные с кибератаками.

Аномалии, криптографические методы и машинное обучение — это некоторые из методов защиты от атак на автомобильные сети. В этой работе мы рассматриваем задачу мультиклассовой классификации и используем машинное обучение для классификации атак. Мы оценили эффективность моделей XGBoost

и логистической регрессии на основе реальных данных из набора данных для хакинга автомобилей.

Цель данной работы – разработка и тестирование моделей машинного обучения для автоматического обнаружения атак в автомобильной сети CAN. Мы сравним два алгоритма – XGBoost и логистическую регрессию – и оценим их способность классифицировать нормальное и атакующее поведение в сети автомобиля.

Обзор литературы

В работе (Chevalier et all., 2021) рассматривается проблема обнаружения кибератак в современных транспортных средствах. Авторы отмечают, что развитие интеллектуальных транспортных систем и широкое использование бортовых сетей передачи данных (в частности, CAN-шины) повышает уязвимость автомобилей к кибератакам. Традиционные методы защиты, основанные на сигнатурном анализе, не всегда способны выявлять новые или модифицированные типы атак, что требует внедрения более гибких и адаптивных подходов. В исследовании предлагаются два метода обнаружения аномалий в автомобильной сети. Первый основан на использовании характеризующих функций (Characteristic Functions), позволяющих выделять статистические и структурные особенности потока сообщений и выявлять отклонения от нормального поведения системы. Второй метод реализован с применением искусственных нейронных сетей (ANN), обучаемых на данных нормального и атакующего трафика. Дополнительно применяется визуальный анализ для интерпретации результатов и оценки характера выявленных аномалий. Особое внимание уделяется сравнению методов по показателям точности обнаружения и вычислительной эффективности. Экспериментальные результаты показывают, что подход на основе характеризующих функций демонстрирует сопоставимую с нейронными сетями точность, при этом значительно превосходит их по скорости обработки данных и требованиям к вычислительным ресурсам. Это делает его более подходящим для внедрения во встроенные автомобильные системы с ограниченными аппаратными возможностями. Работа подтверждает перспективность использования интеллектуальных методов анализа данных для обеспечения кибербезопасности транспортных средств и предлагает практико-ориентированное решение, пригодное для применения в реальных автомобильных инфраструктурах.

В работе (Sharma et all., 2024) авторы исследуют возможности использования методов контролируемого машинного обучения для выявления киберугроз в реальном времени на основе данных из датасета CTU-13, содержащего сетевой трафик с ботнет-атаками. Работа направлена на повышение точности и скорости обнаружения вредоносной активности в сетях. Основное внимание уделено сравнительному анализу таких алгоритмов, как Random Forest, SVM и Gradient Boosting. Оценка эффективности проводится по метрикам точности (accuracy), полноты (recall), F1-меры и времени обработки. Авторы приходят к выводу, что модели Random Forest и Gradient Boosting демонстрируют наилучший

баланс между точностью классификации и производительностью при обработке сетевого трафика. Кроме того, в статье подчеркивается значимость использования реальных наборов данных (как CTU-13) для построения надежных систем обнаружения угроз. Работа представляет интерес для исследователей и практиков в области информационной безопасности, поскольку предлагает обоснованные подходы к применению машинного обучения для задач киберзащиты в условиях ограниченного времени и ресурсов.

В работе (Jabia Nzi et all., 2022) представлено сравнение алгоритмов обнаружения сетевых атак типа DDoS-атак (отказ в обслуживании) для различных сервисов хранения, обработки и передачи данных через Интернет. Особое внимание уделяется применению алгоритмов машинного обучения, таких как гауссовская смешанная модель для максимизации ожиданий (GMM-EM), линейная регрессия (LR), SVM (машина опорных векторов) (с линейным, RBF (радиальная базисная функция) или полиномиальные ядра), алгоритмы дерева решений (Decision Tree), наивный Байеса (Naive Bayes) и рандом Forest (Rand om Forest) для обнаружения такого типа атак. В конце статьи оцениваются перечисленные выше алгоритмы машинного обучения и тщательно сравнивается их производительность. Все экспериментальные результаты показывают, что более 99,7 % двух видов DOS-атак успешно обнаруживаются. Этот подход не снижает производительность и может быть легко распространен на более широкие DOS-атаки.

Исследование (Barthwal et all., 2023) посвящено разработке объяснимой глубокой модели обнаружения вторжений (XAI-based Deep Learning IDS) для интеллектуальных транспортных сетей, основанных на Интернете вещей (IoT). Авторы подчеркивают растущую уязвимость таких систем из-за их связности и сложности, что требует интеграции кибербезопасности на уровне сети. В работе предложен гибридный подход, сочетающий методы глубокого обучения с механизмами объяснимости, что позволяет не только выявлять аномалии с высокой точностью, но и интерпретировать решения модели. Используются современные архитектуры нейронных сетей, включая CNN и LSTM, а также объяснительные методы, такие как LIME и SHAP, для анализа влияния признаков. Эксперименты на наборах данных IoT показали повышение точности обнаружения атак и улучшение прозрачности модели. Исследование делает вклад в развитие надежных, интерпретируемых систем безопасности для «умного» транспорта будущего.

В работе (Wagh et all., 2024) рассматривается применение алгоритмов машинного обучения для обнаружения кибератак и сетевых вторжений в современных коммуникационных системах. Авторы отмечают, что с ростом объема сетевых данных и увеличением сложности атак традиционные методы защиты становятся недостаточно эффективными. В исследовании предлагается модель, основанная на машинном обучении, использующая алгоритмы классификации, такие как Random Forest, Support Vector Machine (SVM) и

Decision Tree, для идентификации аномального поведения в сетевом трафике. Особое внимание уделяется сравнению эффективности различных алгоритмов по показателям точности, полноты и времени обучения. Результаты экспериментов показывают, что методы на основе Random Forest демонстрируют наилучшие результаты при обнаружении известных и неизвестных атак. Работа подчеркивает значимость машинного обучения в повышении безопасности сетей и формирует основу для дальнейшего внедрения интеллектуальных систем защиты в реальных инфраструктурах.

Работа (SaiKiran et all., 2025) посвящена разработке интеллектуального подхода к обнаружению кибератак в компьютерных сетях с использованием технологий машинного обучения. Авторы подчеркивают необходимость автоматизированных и адаптивных систем защиты, способных эффективно противостоять быстро эволюционирующим угрозам. В работе представлена модель, использующая комбинацию алгоритмов классификации, включая K-Nearest Neighbors (KNN), Random Forest и Logistic Regression, для анализа сетевого трафика и выявления аномалий. Особое внимание уделено предварительной обработке данных и выбору релевантных признаков, что позволяет повысить точность классификации атак. Результаты экспериментальных испытаний показали, что предложенный подход обеспечивает высокую точность и надежность при низком уровне ложных срабатываний. Исследование демонстрирует потенциал машинного обучения для создания интеллектуальных, устойчивых к новым угрозам систем кибербезопасности и подчеркивает важность их интеграции в современные сетевые инфраструктуры.

Исследование (Maltseva et all., 2024) посвящено разработке алгоритмов раннего обнаружения кибератак на сети с использованием методов машинного обучения. Авторы отмечают, что традиционные системы обнаружения вторжений часто реагируют с опозданием, что приводит к значительным рискам для информационной безопасности. В исследовании предлагается методология, направленная на выявление признаков атак на ранних стадиях, до того как они могут нанести ущерб инфраструктуре. Для этого используются алгоритмы машинного обучения, включая нейронные сети, деревья решений и метод опорных векторов (SVM), а также методы отбора признаков и оптимизации параметров моделей. Проведённые эксперименты показали, что интеграция нескольких алгоритмов повышает точность прогнозирования и снижает количество ложных тревог (Roman et all., 2018). Работа делает вклад в развитие интеллектуальных систем раннего предупреждения, способных адаптироваться к новым типам атак и обеспечивать более высокий уровень киберзащиты сетевых систем.

В работе (Rahman et all.,.2025) рассматривается применение алгоритмов машинного обучения для мониторинга и обнаружения кибератак в динамически изменяющихся сетевых средах. Авторы подчеркивают, что рост числа киберугроз требует автоматизированных, масштабируемых систем, способных анализировать большие объемы данных в реальном времени. В работе представлены и

сравнительно оценены различные алгоритмы машинного обучения — включая Decision Tree, Random Forest, Naïve Bayes и Support Vector Machine (SVM) — с точки зрения их точности, скорости и устойчивости к шумным данным. Результаты экспериментов показывают, что ансамблевые методы, особенно Random Forest, обеспечивают наилучший баланс между точностью и производительностью. Исследование акцентирует внимание на интеграции ML-подходов в системы сетевого мониторинга и их способности выявлять как известные, так и новые типы атак, что способствует повышению общей эффективности и адаптивности средств киберзащиты.

Анализ рассмотренных исследований показывает, что современные подходы к обнаружению кибератак в сетях, включая транспортные системы, активно развиваются в направлении использования алгоритмов машинного обучения и глубокого обучения. Большинство обзорных работ показывают высокую эффективность методов Random Forest, SVM, Decision Tree и Gradient Boosting, которые обеспечивают высокую точность классификации и позволяют выявлять аномалии в реальном времени. Отдельное внимание уделяется объяснимости моделей и раннему обнаружению угроз, что особенно важно для критических инфраструктур, таких как автомобильные сети CAN и IoT-среды. Исследователи сходятся во мнении, что интеграция интеллектуальных алгоритмов в системы мониторинга значительно повышает уровень защиты и снижает риск успешных атак. При этом остаются актуальными задачи повышения устойчивости моделей к новым типам угроз, оптимизации времени обработки и адаптации решений к ресурсно-ограниченным системам, что определяет направления дальнейших исследований в области кибербезопасности транспортных сетей.

Материалы и методы.

В качестве исходных данных для исследования использовался открытый набор Car Hacking Dataset, содержащий журналы сообщений автомобильной сети CAN (Controller Area Network), записанные как в нормальных условиях функционирования транспортного средства, так и в ситуациях, моделирующих кибератаки различных типов — DoS, Fuzzy, RPM Spoofing и Gear Spoofing. Данный набор данных был выбран благодаря своей реалистичности и репрезентативности, поскольку отражает реальные сценарии функционирования электронных блоков управления (ECU), взаимодействующих между собой посредством шины CAN. Каждый экземпляр данных включает временную метку (Timestamp), идентификатор сообщения (CAN_ID) в шестнадцатеричном формате, длину данных (DLC), байты данных (DATA[0–7]) и флаг состояния (Flag), указывающий, является ли сообщение нормальным (R) или атакующим (T). Анализ данных выявил значительный дисбаланс классов: нормальные сообщения составляют около 85,93 %, атакующие — 14,07 %. Это требует применения специальных методов балансировки данных, таких как oversampling (дублирование меньшего класса) или undersampling (сокращение большего класса). Применение данных методов позволяет избежать смещения модели в сторону преобладающего класса и

обеспечить более справедливое обучение. Таким образом, исходный датасет был предварительно очищен от дубликатов, нормализован и преобразован в формат, пригодный для машинного обучения. Такая подготовка данных является ключевым этапом, обеспечивающим адекватную работу моделей классификации атак и снижение вероятности переобучения.

Для оценки взаимосвязей между признаками, извлечёнными из сообщений CAN, была построена матрица корреляции Пирсона (рис. 1). Анализ показал, что между большинством признаков наблюдается слабая или умеренная корреляция, что свидетельствует о низкой избыточности данных и обоснованности их включения в модель. Наиболее выраженные взаимосвязи зафиксированы между признаками DATA[1], DATA[2], DATA[3] и DATA[4], где коэффициенты корреляции достигают значений 0.3–0.4, что указывает на частичную зависимость передаваемых байтов внутри одного пакета. Отрицательная корреляция между CAN_ID и DATA[7] (около –0.21) демонстрирует, что данные байты характеризуют различные аспекты сетевой активности.

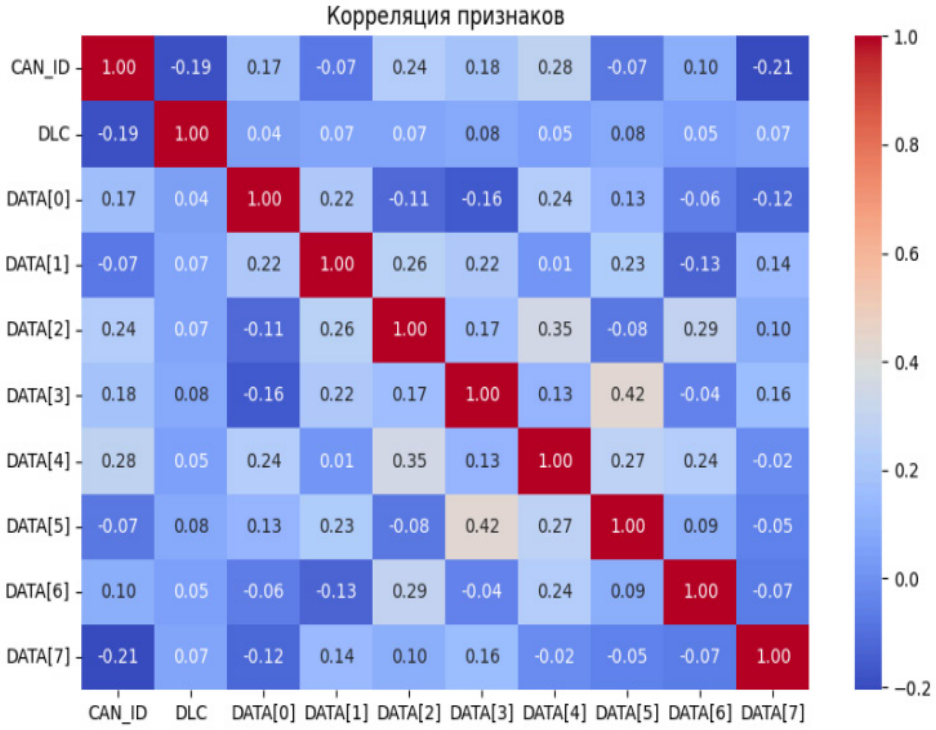

Рис. 1. Матрица корреляции признаков для данных автомобильной сети CAN

Такое распределение корреляций подтверждает целесообразность использования нелинейных методов обучения, таких как XGBoost, способных выявлять сложные взаимодействия между признаками, не ограничиваясь линейными зависимостями. При этом низкая взаимная корреляция между большинством признаков минимизирует риск мультиколлинеарности, что

положительно сказывается на стабильности и обобщающей способности модели. Таким образом, анализ корреляции подтверждает корректность предварительного отбора признаков и обеспечивает дополнительное обоснование применённого подхода к построению модели обнаружения кибератак.

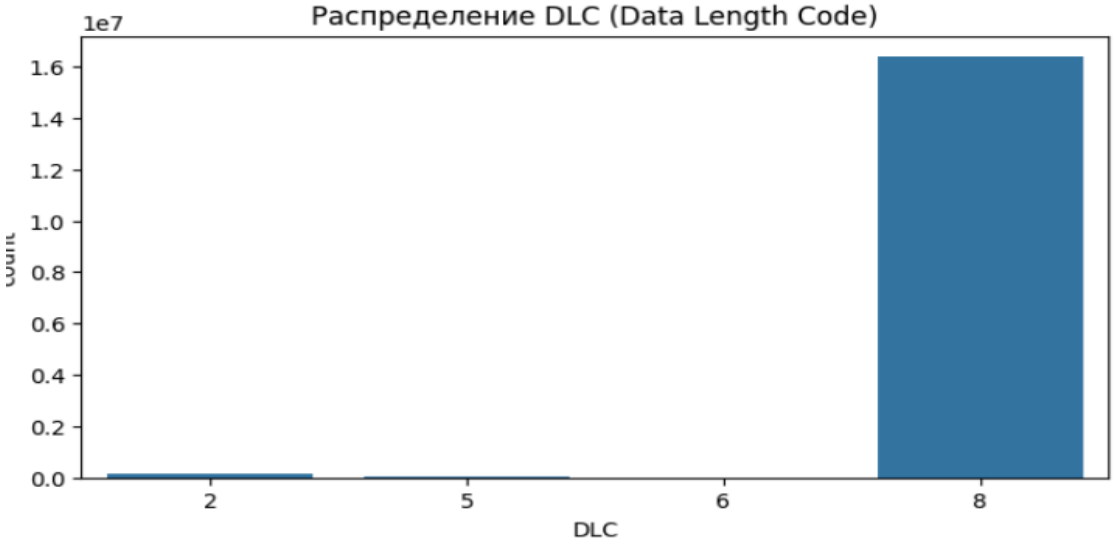

Рис. 2. Распределение длины пакета данных (DLC) в CAN-сообщениях

На рисунке 2 представлено распределение длины пакета данных (Data Length Code, DLC**)** для всех сообщений автомобильной сети CAN. Анализ показал, что подавляющее большинство пакетов имеют длину 8 байт**,** что соответствует максимально возможному объёму данных в одном сообщении CAN и отражает типичную структуру обмена информацией между электронными блоками управления (ECU). Такая закономерность объясняется особенностями формирования сообщений в автомобильных сетях, где наиболее информативные данные (например, обороты двигателя, положение педали акселератора или параметры трансмиссии) передаются в полном формате 8-байтных кадров.

Сообщения с меньшей длиной (1–7 байт) встречаются крайне редко и, как правило, связаны с диагностическими или служебными сигналами. Это распределение подтверждает однородность структуры трафика и объясняет низкую вариативность признака DLC в дальнейших моделях машинного обучения. В контексте задачи обнаружения атак данная особенность указывает на то, что длина сообщения не является значимым предиктором для классификации аномалий, однако может использоваться в сочетании с другими признаками (например, CAN_ID или значениями DATA[0–7]) для комплексного анализа сетевого поведения. Таким образом, анализ DLC подтверждает стабильность сетевой структуры и отсутствие явных аномалий в параметре длины пакета.

Перед обучением модели данные проходят этап предобработки, который

включает несколько обязательных шагов. В первую очередь выполняется очистка данных: удаляются дубликаты, а также обрабатываются пропущенные значения, чтобы избежать искажения результатов обучения. Далее проводится преобразование признаков — идентификаторы CAN ID переводятся в числовой формат, а числовые характеристики нормализуются для обеспечения корректной работы алгоритмов машинного обучения. После этого набор данных разделяется на обучающую и тестовую выборки в стандартном соотношении 80/20. Дополнительно применяется балансировка классов с использованием методов увеличения или уменьшения выборки, что позволяет повысить качество классификации и снизить влияние дисбаланса данных.

Для решения задачи классификации атак используются два алгоритма машинного обучения. В качестве основного метода рассматривается XGBoost — эффективный бустинговый алгоритм, хорошо справляющийся с задачами классификации и устойчивый к дисбалансу классов. В качестве базовой модели применяется логистическая регрессия, отличающаяся простотой и стабильной работой на линейно разделимых данных. Обучение моделей проводится на предварительно сбалансированных данных с применением кросс-валидации, что позволяет повысить надежность и обобщающую способность полученных результатов.

Результаты и обсуждение.

Матрица ошибок (Confusion Matrix) показала, что модель XGBoost более точно классифицирует атакующие сообщения по сравнению с логистической регрессией. Последняя в ряде случаев ошибочно относила атакующие сообщения к нормальному поведению системы. Для оценки качества классификации модели XGBoost была построена матрица ошибок, позволяющая проанализировать распределение правильно и ошибочно классифицированных объектов по каждому классу (рис.3).

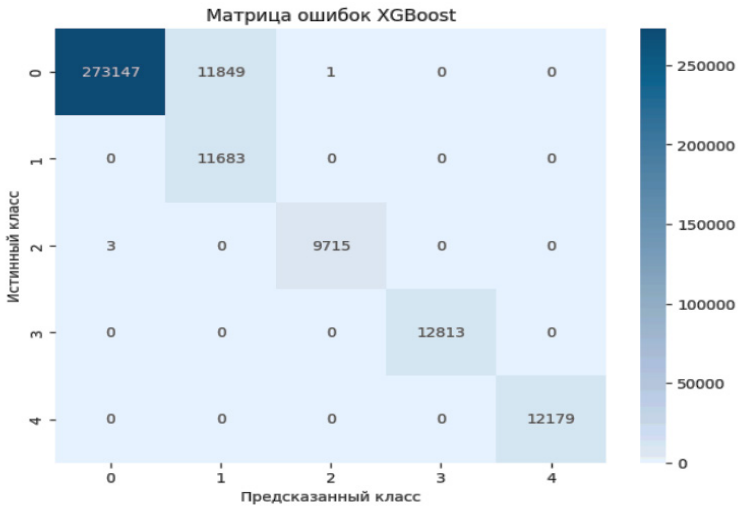

Рис.3. Матрица ошибок модели XGBoost

Для оценки точности классификации и выявления характера ошибок при использовании модели логистической регрессии была построена матрица ошибок, позволяющая определить распределение верно и ошибочно отнесённых объектов по классам (Рис. 4).

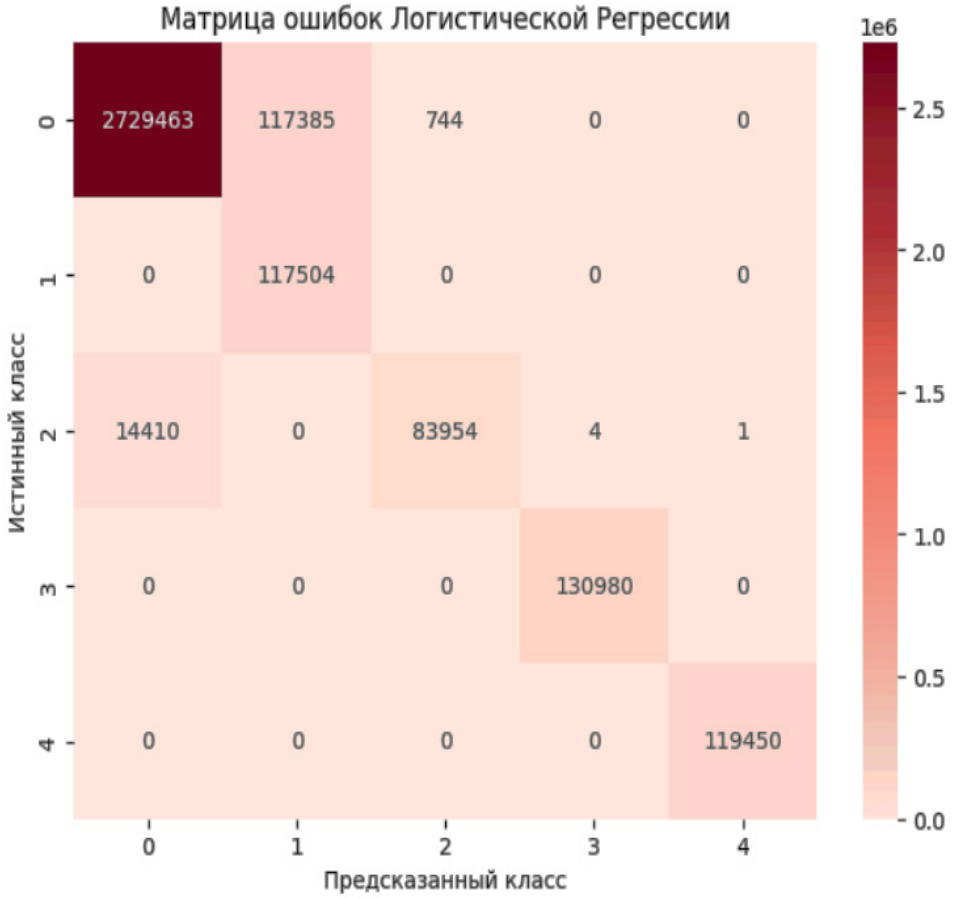

Рис. 4. Матрица ошибок модели логистической регрессии

График важности признаков, полученный на основе модели XGBoost, показал, что наибольший вклад в процесс классификации вносят параметры DATA[7] и DATA[5]**.** Высокая значимость указанных признаков свидетельствует о том, что именно они в наибольшей степени отражают характерные особенности инжектируемых атак и оказывают существенное влияние на принятие моделью классификационного решения.

В случае использования логистической регрессии наиболее значимыми оказались признаки DATA[6] и DATA[5]**.** Наличие общего значимого параметра DATA[5] для обеих моделей указывает на его устойчивую информативность и позволяет рассматривать данный признак как один из ключевых факторов, описывающих закономерности инжектируемых атак. Отличия в ранжировании остальных

признаков объясняются различиями в принципах работы алгоритмов: XGBoost выявляет сложные нелинейные зависимости и взаимодействия между признаками, тогда как логистическая регрессия ориентирована преимущественно на линейные взаимосвязи.

Полученные результаты подтверждают, что анализ важности признаков позволяет выявить скрытые зависимости в структуре данных и повысить интерпретируемость моделей машинного обучения при решении задач обнаружения инжектируемых атак. Для интерпретации результатов классификации и выявления наиболее информативных признаков был выполнен анализ их важности для моделей XGBoost и логистической регрессии (рис.5). Оценка вклада отдельных признаков позволяет определить, какие параметры в наибольшей степени влияют на процесс принятия классификационного решения и отражают характерные особенности инжектируемых атак.

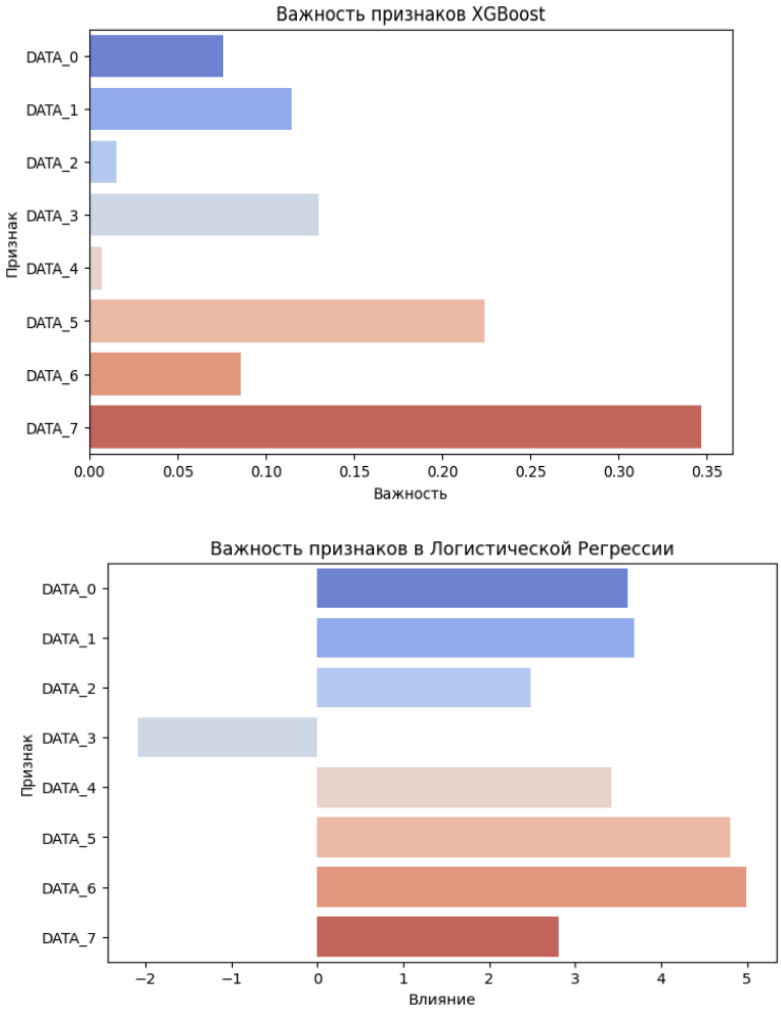

Рис. 5. Важность признаков в модели XGBoost

Проведённый анализ результатов классификации показал, что используемые модели демонстрируют высокую точность при распознавании объектов, относящихся к классам 0, 2, 3 и 4. Для указанных классов наблюдается преобладание корректных предсказаний, что свидетельствует о достаточной информативности используемых признаков и устойчивости моделей в данных категориях. В то же время классификация объектов класса 1 сопровождается снижением точности. Это обусловлено тем, что характеристики данного класса в значительной степени пересекаются с признаковым пространством других классов, прежде всего класса **0**. В результате объекты класса 1 часто ошибочно относятся к классу 0, что приводит к увеличению числа ошибок и снижению показателей точности и полноты для данного класса. Полученные результаты указывают на необходимость дополнительной дифференциации класса 1, в том числе за счёт расширения набора признаков, применения методов балансировки классов либо использования более сложных моделей, способных лучше разделять близкие по характеристикам классы.

Для более детальной оценки качества работы моделей была проанализирована точность предсказаний по каждому классу отдельно. Такой анализ позволяет выявить классы, для которых модели демонстрируют устойчивые результаты, а также определить категории, вызывающие наибольшие затруднения при классификации (рис.6).

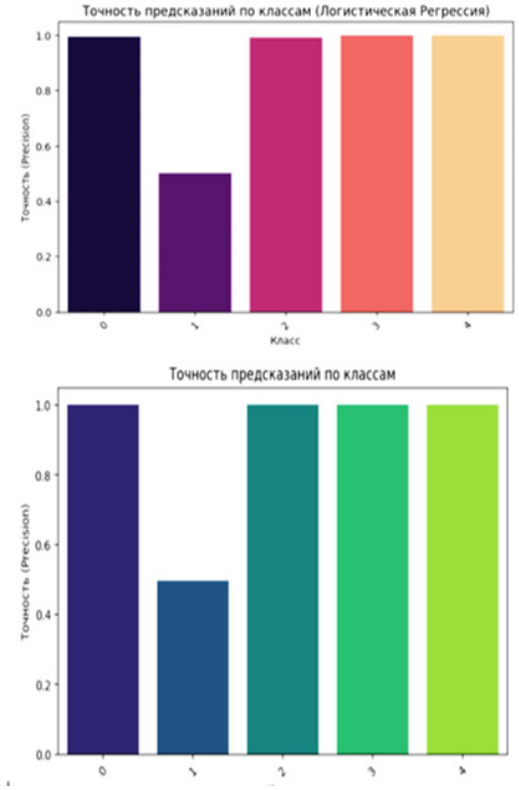

Рис. 6. Точность предсказаний по классам для модели логистической регрессии

Полученные экспериментальные результаты свидетельствуют о высокой эффективности применения методов машинного обучения для обнаружения кибератак в автомобильных сетях CAN. Сравнительный анализ показал, что модель XGBoost превосходит логистическую регрессию по точности классификации и способности выявлять атакующее поведение, что объясняется ее способностью учитывать нелинейные зависимости и сложные взаимодействия между признаками.

Анализ матриц ошибок продемонстрировал, что обе модели уверенно классифицируют большинство классов атак, однако наибольшие трудности возникают при распознавании класса 1. Это связано с пересечением его характеристик с признаковым пространством нормального трафика, что приводит к ошибочной классификации части атак как легитимных сообщений. Данный факт указывает на необходимость дальнейшего расширения признакового пространства или применения более сложных ансамблевых и гибридных моделей.

Дополнительный анализ важности признаков показал, что параметры DATA[5], DATA[6] и DATA[7] играют ключевую роль в выявлении инжектируемых атак. Совпадение наиболее значимых признаков для разных моделей подтверждает их устойчивую информативность и практическую значимость для задач мониторинга CAN-трафика. В целом результаты согласуются с выводами современных исследований и подтверждают перспективность использования XGBoost для задач кибербезопасности транспортных систем.

Заключение.

В ходе проведённого исследования была решена актуальная задача обнаружения и классификации кибератак в автомобильных сетях CAN на основе методов машинного обучения. Рост цифровизации транспортных средств и увеличение числа электронных блоков управления существенно повышают требования к обеспечению кибербезопасности бортовых сетей, что делает разработку интеллектуальных систем мониторинга особенно востребованной. В работе использован открытый набор данных Car Hacking Dataset, содержащий реальные журналы CAN-сообщений как в штатном режиме, так и при моделировании атак типов DoS, Fuzzy, RPM Spoofing и Gear Spoofing. Проведена комплексная предобработка данных, включающая очистку, нормализацию, анализ корреляции признаков и балансировку классов, что позволило повысить устойчивость моделей и снизить влияние дисбаланса выборки. Для решения задачи мультиклассовой классификации были реализованы и сравнены два алгоритма — логистическая регрессия и XGBoost. Результаты экспериментов показали, что модель XGBoost обеспечивает более высокую точность и устойчивость при распознавании атакующих классов по сравнению с линейной моделью, что обусловлено её способностью учитывать сложные нелинейные зависимости между признаками CAN-сообщений. Анализ матриц ошибок позволил выявить классы, вызывающие наибольшие трудности при классификации, что указывает на необходимость дальнейшего расширения признакового пространства и

совершенствования методов разделения близких по характеристикам классов. Дополнительный анализ важности признаков показал, что наибольший вклад в процесс классификации вносят отдельные байты данных CAN-сообщений, что подтверждает их информативность при выявлении инжектируемых атак и повышает интерпретируемость построенных моделей. Практическая значимость работы заключается в возможности интеграции предложенного подхода в интеллектуальные системы обнаружения вторжений, функционирующие в условиях ограниченных вычислительных ресурсов автомобильных платформ. Полученные результаты подтверждают эффективность применения методов машинного обучения для повышения уровня кибербезопасности транспортных сетей и создают основу для дальнейших исследований, направленных на адаптацию моделей к работе в режиме реального времени, расширение спектра анализируемых атак и разработку более устойчивых и масштабируемых решений для защиты современных транспортных средств. Перспективным направлением дальнейших исследований является применение более сложных архитектур машинного обучения, включая ансамблевые методы и модели глубокого обучения (например, нейронные сети CNN и LSTM), что позволит повысить точность классификации трудноразделимых классов. Кроме того, целесообразно провести оценку вычислительной сложности и времени обработки предложенного подхода, что позволит определить его применимость в системах реального времени и в условиях ограниченных вычислительных ресурсов автомобильных платформ.

REFERENCES

Barthwal A., & Raheja S. (2023). An explainable deep learning intrusion detection in IoT-enabled transportation networks // Proceedings of the International Conference on Artificial Intelligence and Computing Communication Technologies. Pp. 310–316. 10.1109/ICAICCIT60255.2023.10466149.

Cil A. E., Yildiz K., & Buldu A. (2021). Detection of DDoS attacks with feed forward based deep neural network model // Expert Systems with Applications. P.169. Article 114520. 10.1016/j.eswa.2020.114520.

Chevalier Y., Fenzl F., Kolomeets M., Rieke R., Chechulin A., & Kraus C. (2021). Cyberattack Detection in Vehicles using Characteristic Functions, Artificial Neural Networks, and Visual Analysis. — Informatics and Automation. Vol. 20(4). — Pp. 845–868. 10.15622/ia.20.4.4.

Jabia Nzi J. M., & Safaryan O. A. (2022). Investigation of DDoS attack detection using machine learning // The Young Researcher of the Don. No. 6(39). URL: https://cyberleninka.ru/article/n/issledovanie-obnaruzheniya-ddos-atak-s-ispolzovaniem-mashinnogo-obucheniya [in Russ.].

Maltseva I., Chernysh Yu., & Protsyuk Yu. (2024). Analysis of algorithms for early detection of cyber attacks on networks using machine learning // Communication, Informatization, and Cybersecurity Systems and Technologies. Vol. 1(6). Pp. 105–115. 10.58254/viti.6.2024.08.105 [in Ukr.].

Sharma A., & Babbar H. (2024). Detecting cyber threats in real time: A supervised learning perspective on the CTU-13 dataset // Proceedings of the IEEE International Conference for Innovation in Technology (INOCET). 10.1109/INCET61516.2024.10593100.

SaiKiran N., & Jagadeesh K.A. (2025). An intelligent approach to cyber-attack detection in networks using machine learning techniques // *International Journal of Research and Innovation in Applied Science*. Pp. 1351–1358. 10.51584/ijrias.2025.100800117.

Roman R., Lopez J., & Mambo M. (2018). Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. Future Generation Computer Systems. Vol. 78(2). Pp. 680–698. 10.1016/j.future.2016.11.009.

Rahman M. T., & Rahman Md. K. (2025). Machine learning algorithms for monitoring and detecting cyber attacks. // Proceedings of the International Conference on Machine Learning Applications. 2025. Pp. 1–6. 10.1109/mac64480.2025.11140542.

Wagh A., Pawar R., Wable N., Wandhekar S., & Dighe M. S. (2024). Detection of cyber attacks and network attacks using machine learning algorithms // *International Journal of Advanced Research in Science, Communication and Technology*. 10.48175/ijarsct-18161.

Xiao Y., Jia Y., Liu C., Cheng X., Yu J., & Lv W. (2019). Edge computing security: State of the art and challenges. // Proceedings of the // IEEE. 2019. Vol. 107(8). Pp. 1608–1631. URL: https://www.semanticscholar.org/paper/Edge-Computing-Security%3A-State-of-the-Art-and-Xiao-Jia/5f5ccd80381ca3593f9fc651844ed506894cbaf7

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

Собственник:
АО «Международный университет информационных технологий» (Казахстан, Алматы)

Главный редактор:
Колесникова Катерина Викторовна

Ответственный редактор:
Мрзабаева Раушан Жалиевна

Компьютерная верстка:
Калабай Замзагуль Ертугановна

Сайт журнала: https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Подписано в печать 30.03.2026.
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание АО "Международного университета информационных технологий"