Научная статья
DOI 10.66424/2071-8217-2026-2-7
УДК 004.032.26

АУГМЕНТАЦИЯ ТРАФИКА ИНТЕРНЕТА ВЕЩЕЙ С ИСПОЛЬЗОВАНИЕМ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ

В. В. Платонов[1*], Д. А. Скиба[2]

[1]Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия
[2]АО «ИнфоТеКС», Москва, Россия

✉ *plato@ibks.spbstu.ru

ДЛЯ ЦИТИРОВАНИЯ

Платонов В. В., Скиба Д. А. Аугментация трафика Интернета вещей с использованием генеративно-состязательных сетей // Проблемы информационной безопасности. Компьютерные системы. 2026. № 2. С. 82–91.
DOI: 10.66424/2071-8217-2026-2-7

ПОСТУПИЛА 09.02.2026
ПРИНЯТА 04.05.2026
ОПУБЛИКОВАНА 15.06.2026

Издатель: Санкт-Петербургский политехнический университет Петра Великого

АННОТАЦИЯ

Исследуется проблема критического дисбаланса классов в системах обнаружения вторжений (IDS) для сетей Интернета вещей (IoT). Проведено сравнительное исследование результативности различных методов аугментации данных: пяти архитектур генеративно-состязательных сетей (CopulaGAN, CTGAN, CTAB-GAN+ и модификаций MC-WGAN-GP и TMG-GAN) в сопоставлении с традиционными подходами (SMOTE, случайное пересемплирование). Показано, что применение аугментации (как этапа подготовки данных) позволяет восстановить работоспособность классификатора LightGBM в сценариях с критическим дисбалансом, увеличивая показатель F1-macro с 0,03 до 0,81.

КЛЮЧЕВЫЕ СЛОВА

Аугментация данных, генеративно-состязательные сети, дефицит данных, система обнаружения вторжений, Интернет вещей

Original article
DOI 10.66424/2071-8217-2026-2-7

IOT DATA AUGMENTATION USING GENERATIVE ADVERSARIAL NETWORKS

V. V. Platonov[1*], D. A. Skiba[2]

[1]Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
[2]JSC “InfoTeX”, Moscow, Russia

✉ *plato@ibks.spbstu.ru

FOR CITATION

Platonov V. V., Skiba D. A. IoT data augmentation using generative adversarial networks. *Problems of information security. Computer systems*. 2026. No. 2, pp. 82–91.
DOI: 10.66424/2071-8217-2026-2-7
(In Russian)

ABSTRACT

The article investigates the problem of critical class imbalance in intrusion detection systems (IDS) for Internet of Things (IoT) networks. A comparative study of data augmentation methods was conducted, evaluating five generative adversarial network (GAN) architectures (CopulaGAN, CTGAN, CTAB-GAN+ and modified versions of MC-WGAN-GP and TMG-GAN) against traditional approaches (SMOTE, random oversampling). The study shows that data augmentation (as a

RECEIVED 09.02.2026
ACCEPTED 04.05.2026
PUBLICATION 15.06.2026

data preprocessing stage) enables the restoration of the LightGBM classifier's performance in critical imbalance scenarios, increasing the F1-macro score from 0.03 to 0.81.

KEYWORDS

Data augmentation, generative adversarial networks, data deficiency, intrusion detection system, Internet of Things

1. ВВЕДЕНИЕ

В эпоху цифровизации, когда сети Интернета вещей (Internet of Things – IoT) проникают во все сферы жизни – от промышленных систем и умных городов до бытовых устройств, – объем генерируемого трафика достигает триллионов пакетов ежедневно, создавая новые вызовы для обеспечения кибербезопасности. По оценкам аналитиков, в 2025 г. количество подключенных IoT-устройств превысит 20 млрд с прогнозируемым ростом на 14% ежегодно [1]. Это неизбежно ведет к эскалации угроз: ежедневно фиксируется около 820 тыс. атак [2] на IoT-инфраструктуру, включая DDoS, ботнеты и сетевую разведку. Согласно отчету Nozomi Networks Labs, в первой половине 2025 г. вредоносная активность в отношении критической инфраструктуры значительно возросла, при этом преобладают brute-force атаки и эксплуатация устаревших уязвимостей [3].

Эффективная защита таких сетей невозможна без современных систем обнаружения вторжений (Intrusion Detection System – IDS), использующих алгоритмы машинного обучения для идентификации аномального поведения. Однако на практике разработчики IDS сталкиваются с фундаментальной проблемой дефицита и дисбаланса данных (class imbalance). В реальных условиях эксплуатации доля атакующего трафика ничтожно мала по сравнению с фоновым (нормальным трафиком), что приводит к дисбалансу классов в обучающих выборках и, как следствие, к деградации классификаторов – модели склонны игнорировать минорные классы, что недопустимо для систем безопасности.

Большинство открытых наборов данных для обучения IDS, таких как актуальный CIC-IoT-2023 [4, 5], характеризуются высокой плотностью вредоносного трафика, что упрощает задачу обучения, но не соответствует реальным условиям функционирования IoT-сетей. Для имитации эксплуатации системы в реальной среде необходимо создание условий критического дисбаланса путем искусственного увеличения минорных классов. В таких сценариях традиционные методы борьбы с дисбалансом, такие как случайное дублирование (ROS) или синтетическая генерация на основе ближайших соседей (SMOTE [6]), часто оказываются недостаточно эффективными для восстановления сложной структуры NetFlow-данных.

Перспективным решением является применение генеративно-состязательных сетей (Generative Adversarial Networks – GAN), способных моделировать многомерные распределения признаков. Приведены результаты исследования эффективности различных архитектур GAN (включая CopulaGAN, CTGAN [7] и авторские модификации) для синтеза табличных данных IoT-трафика в условиях моделируемого критического дефицита. Представлен сравнительный анализ влияния аугментации на качество классификации при различных параметрах дисбаланса, объема данных и интенсивности синтеза.

Методы борьбы с дисбалансом можно разделить на алгоритмические (настройка весов и гиперпараметров) и методы уровня данных (пересемплирование). В данной работе акцент сделан на модификации данных, так как этот подход является универсальным и позволяет подготовить качественный обучающий набор, пригодный для использования с любыми архитектурами классификаторов без необходимости их специфической настройки под каждый тип атаки.

2. МЕТОДЫ

В качестве исходного набора данных использовался набор CIC-IoT-2023. В нем данные представлены в виде потоков трафика (формат netflow). Набор данных содержит восемь классов: семь вредоносных и один легитимный.

Параметры аугментации. В качестве параметров, контролирующих условия аугментации, выбраны:

1. Объем данных N_B – количество экземпляров в легитимном классе.

2. Степень дисбаланса (ir – imbalance ratio):

$$ir = \frac{N_M}{N_B},$$

где N_M – количество экземпляров в одном вредоносном классе обучающего (несбалансированного) набора данных.

3. Интенсивность аугментации (tr – target ratio):

$$tr = \frac{\tilde{N}_M}{N_B},$$

где $\tilde{N}_M$ – количество экземпляров в одном вредоносном классе после аугментации (в аугментированном наборе данных).

Наборы параметров аугментации разделены на четыре группы:

1. Основной сценарий. Сильный дефицит и дисбаланс данных. Параметры:

• N_B (объем данных) ∈ {5000, 10000, 20000};

• ir (степень дисбаланса) ∈ {0,0075; 0,015};

• tr (интенсивность аугментации) ∈ {0,05; 0,1; 0,15; 0,2}.

2. Высокая интенсивность аугментации ($tr > 0{,}25$).

3. Слабый дисбаланс ($ir > 0{,}05$).

4. Большой объем данных ($N_B > 30000$).

Наборы данных. Для проведения экспериментов сформированы три выборки (без учета наборов данных для аугментации):

1. Обучающий (несбалансированный) набор данных. Используется для обучения моделей аугментации, а также при классификации для получения «базового» результата в условиях дисбаланса.

2. Дополнительный набор данных. Содержит реальные данные в таком же количестве, в котором будут генерироваться синтетические данные. Используется для сравнения прироста качества классификации при добавлении синтетических и реальных данных в несбалансированный набор.

3. Тестовый набор данных. Предназначен для сбора метрик результатов классификации.

Схема формирования наборов данных представлена на рис. 1.

Методы аугментации. Использовались следующие методы аугментации:

• традиционные: SMOTE, SMOTE-Tomek, случайное пересемлирование, GaussianCopula;

• GAN: CTGAN [7], CTAB-GAN+ [8], MC-WGAN-GP [9], TMG-GAN [10], CopulaGAN [7];

• TVAE [7], используется в качестве генеративной альтернативы GAN.

Использованы реализации CTGAN, TVAE и CopulaGAN из библиотеки SDV [11–13].

Изначально в группу традиционных методов аугментации планировалось вклю-

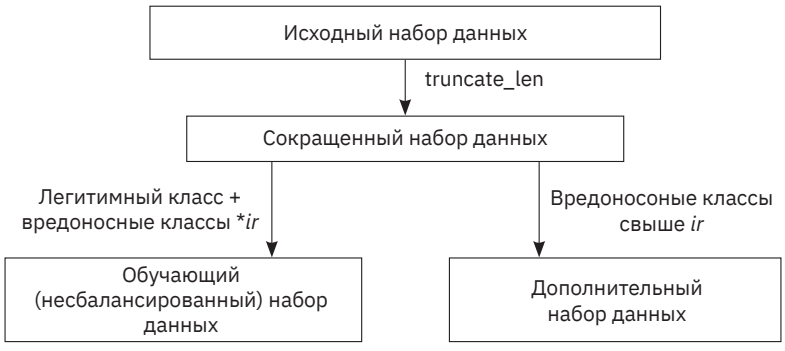

Рис. 1 | Схема формирования наборов данных

Fig. 1 | Data set formation scheme

чить также алгоритм ADASYN (Adaptive Synthetic Sampling). Однако в ходе экспериментов выявлена его принципиальная неприменимость к условиям, наиболее характерным для реальных систем обнаружения вторжений в сетях Интернета вещей.

Модификации моделей. Оригинальные архитектуры MC-WGAN-GP и TMG-GAN обладают узкой специализацией: первая ориентирована исключительно на категориальные признаки, вторая – на непрерывные численные величины. Для преодоления этого ограничения реализована модификация выходного слоя генератора на основе многопоточной архитектуры (multi-head architecture). В рамках данного подхода для каждого атрибута формируется отдельный выходной слой с семантически зависимой функцией активации: Softmax или Gumbel-Softmax для дискретных категорий и сигмоидальная или тождественная функция для аппроксимации численных значений.

Классификаторы. В качестве классификаторов использовались ансамблевые модели: LightGBM [14], RandomForest [15] и XGBoost [16]. При обучении использовались гиперпараметры по умолчанию, кроме параметра n_estimators, для которого задано значение 200 для каждого классификатора.

Работа реализованного прототип включает три этапа:

1. Подготовка несбалансированного набора данных для обучения (с учетом объема данных N_B и степени дисбаланса ir). Сначала количество экземпляров в каждом классе равно N_B, затем количество вредоносных экземпляров уменьшается до значения $N_B ir$.

2. Аугментация данных различными методами (с учетом интенсивности аугментации tr). После аугментации сохраняется аугментированный набор для каждого метода аугментации.

3. Классификация. Обучение классификаторов проводится: на исходных несбалансированных данных; аугментированных данных (10 наборов данных, один для каждого метода аугментации); дополненных данных (в несбалансированный набор добавлено такое количество экземпляров вредоносных классов, чтобы достичь показателя tr).

Схема программного прототипа представлена на рис. 2.

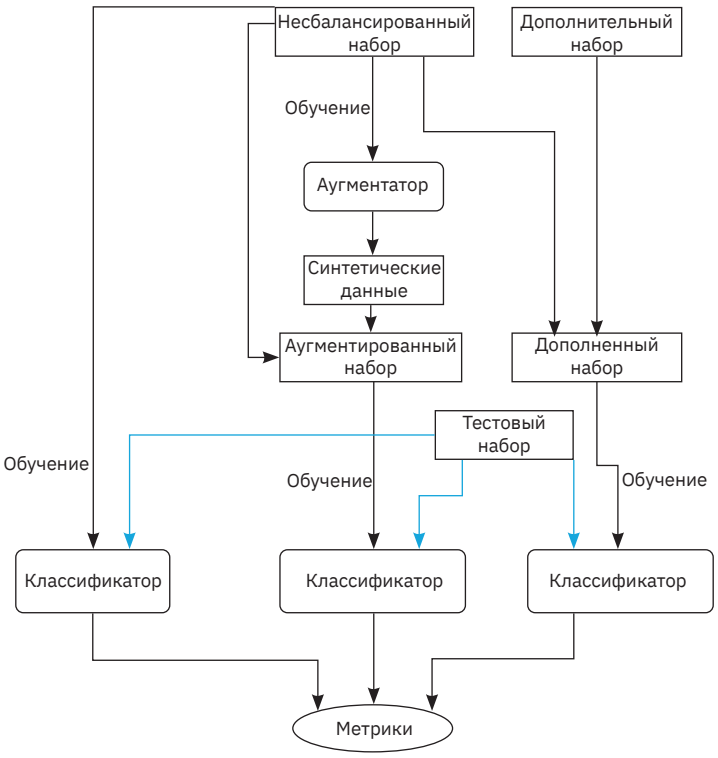

Рис. 2 | Схема разработанного прототипа

Fig. 2 | Diagram of the developed prototype

Метрика. Для оценки качества классификации использовалась макроусредненная F1-мера (далее F1-macro). Она подходит для оценки качества классификации в условиях дисбаланса данных, так как каждый класс имеет одинаковый вес. Прирост F1-macro далее будет указываться в процентах, величина абсолютная.

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

В табл. 1 приведены метрики качества классификации в зависимости от метода аугментации, а также результаты при дополнении данных реальными данными. Сравнительный анализ различных генеративных подходов позволил установить преимущество архитектуры CopulaGAN, которая по ключевым метрикам превзошла альтернативную модель на основе вариационных автокодировщиков (TVAE). Тем не менее в ходе экспериментов зафиксировано, что классические методы пересемплирования, такие как SMOTE, SMOTE-Tomek и случайное пересемплирование, демонстрируют более высокие показатели эффективности, чем CopulaGAN.

Данный результат находит теоретическое обоснование в специфике обучения глубоких нейронных сетей: для адекватной настройки параметров GAN-моделей требуется значительный объем репрезентативной выборки минорных классов. В то же время алгоритм SMOTE, основанный на линейной интерполяции признакового пространства, сохраняет устойчивость в условиях экстремального дефицита данных. Относительно низкие показатели GAN-архитектур на табличных данных формата NetFlow объясняются тем, что структура последних зачастую характеризуется менее сложными нелинейными зависимостями по сравнению с изображениями или текстами. Это подтверждает гипотезу

Таблица 1 | Сравнение методов аугментации

Table 1 | Comparison of augmentation methods

Метод	Средний ΔF1-macro, %	Медиана ΔF1-macro, %	Win-rate (улучшение), %	Максимум, %	Минимум, %
Дополненные данные	+14,0	+11,5	100	+84,2	+0,8
SMOTE	+5,6	+2,6	92	+80,8	–2,7
SMOTETomek	+5,5	+2,3	93	+80,4	–2,7
Случайное пересемплирование	+4,9	+2,3	85	+81,2	–10,2
CopulaGAN	+4,2	+1,1	72	+79,3	–5,5
GaussianCopula	+3,9	+0,9	76	+80,2	–3,3
CTGAN	+2,8	+0,2	54	+77,7	–40,2
TVAE	+2,5	+0,2	54	+80,2	–9,2
MCWGANGP	+1,6	–0,9	23	+75,0	–6,6
CTABPlus	+1,4	–1,0	21	+74,5	–7,0
TMGGAN	+1,2	–0,7	29	+77,7	–13,7

Примечания: Win-rate – это доля экспериментов, в которых модель, обученная на аугментированных данным методом данных, превзошла по метрике F1-macro контрольную модель, обученную на исходных (несбалансированных) данных. Максимум – максимальный прирост F1-macro среди всех экспериментов. Минимум – минимальный прирост F1-macro.

о том, что превосходство состязательных сетей над статистическими методами проявляется лишь при достижении определенного порога объема и сложности входных данных.

Контрольный эксперимент с добавлением реальных данных показал предсказуемо высокий результат (ΔF1-macro = 14 %), превзошедший все методы синтеза. Однако сложность масштабирования стендов и трудоемкость экспертной разметки трафика в реальных IoT-инфраструктурах делают методы генеративной аугментации (в частности, CopulaGAN) более приоритетными для оперативного обучения IDS в условиях априорного дефицита информации об атаках.

Анализ эффективности аугментации в зависимости от условий эксперимента (табл. 2) позволил выявить ряд ключевых закономерностей. В качестве репрезентативной модели для оценки влияния параметров выбран CopulaGAN, как показавший наиболее стабильные результаты среди методов аугментации с использованием GAN.

На основе полученных данных (табл. 2) можно сделать следующие выводы:

1. Влияние интенсивности синтеза. Сравнение основного сценария (строка 1) и сценария с высокой интенсивностью аугментации (строка 2) показывает, что избыточная генерация синтетических данных ($tr > 0{,}25$) приводит к снижению эффективности классификации. Прирост F1-macro падает с 4,8 до 2,2 %, что свидетельствует о накоплении «статистического шума» и переобучении моделей на искусственных паттернах.

2. Порог целесообразности. В условиях слабого дисбаланса (строка 3, $tr > 0{,}05$) применение аугментации нецелесообразно: зафиксировано незначительное снижение качества (–0,1 %). Это подтверждает гипотезу о том, что современные ансамблевые классификаторы обладают достаточной внутренней устойчивостью к умеренному дисбалансу классов.

3. Эффект объема данных. Особый интерес представляет сравнение групп 1 и 4. Увеличение объема исходной несбалансированной выборки само по себе не ведет к росту качества (показатель в столбце «Дисбаланс» остается на уровне 0,7). Однако именно на больших объемах данных аугментация демонстрирует максимальную эффективность (ΔF1-macro = +9,1 %). Это объясняется тем, что увеличение обучающей выборки для генератора позволяет GAN-модели точнее аппроксимировать распределение признаков, минимизируя ошибки при синтезе минорных классов.

Оценка чувствительности различных алгоритмов машинного обучения к аугментации методом CopulaGAN (табл. 3, 4) выявила значительные различия в их архитектурной устойчивости.

Основные выводы:

1. Random Forest демонстрирует исходную устойчивость к дисбалансу благодаря механизму бэггинга. Дополнительная аугментация не приводит к росту целевой метрики, а в ряде случаев вызывает незначительную деградацию (–1,6 %).

2. Градиентный бустинг (XGBoost, LightGBM): в условиях дефицита данных модели склонны к переобучению, что критично

Таблица 2 | Изменение качества классификации (ΔF1-macro) по группам параметров

Table 2 | Change in classification quality (ΔF1-macro) by parameter groups

Номер строки	Группа параметров	Дисбаланс	Δ CopulaGAN, %
1	Основной сценарий	0,69	+4,8
2	Высокая интенсивность аугментации	0,7	+2,2
3	Слабый дисбаланс	0,84	–0,1
4	Большой объем данных	0,7	+9,1

Таблица 3 | Результаты классификаторов при использовании набора параметров «Основной сценарий»

Table 3 | Classifier results when using the "Main Scenario" parameter set

Классификатор	Дисбаланс	ΔCopulaGAN, %	CopulaGAN
RandomForest	0,69	–0,1	0,69
XGBoost	0,73	+2,4	0,75
LightGBM	0,65	+12,2	0,77

Таблица 4 | Результаты классификаторов при использовании набора параметров «Большой объем данных»

Table 4 | Classifier results when using a set of "Large data volume" parameters

Классификатор	Дисбаланс	ΔCopulaGAN, %	CopulaGAN
RandomForest	0,76	–1,6	0,74
XGBoost	0,79	+3,3	0,82
LightGBM	0,57	+25,7	0,83

Таблица 5 | F1-мера каждого класса после аугментации

Table 5 | F1 is the measure of each class after augmentation

Класс	Дисбаланс	CopulaGAN
Легитимный	0,54	0,62
DDoS	0,76	0,99
DoS	0,83	0,99
Mirai	0,9	0,99
Recon	0,53	0,71
Spoofing	0,52	0,76
BruteForce	0,5	0,69
Web	0,42	0,6

для несбалансированных выборок. До аугментации LightGBM показывает наихудшую устойчивость.

3. Синергетический эффект: применение CopulaGAN наиболее эффективно для LightGBM, обеспечивая максимальный прирост (ΔF1-macro до +25,7 %). После аугментации LightGBM превосходит остальные модели, что делает связку CopulaGAN + LightGBM наиболее эффективным решением для задач IDS в условиях экстремального дисбаланса.

При использовании связки LightGBM и CopulaGAN для наборов параметров «Основной сценарий» и «Большой объем данных» F1-мера каждого класса растет после аугментации (табл. 5).

Наибольшая эффективность предложенного подхода зафиксирована в сценарии с максимальным объемом выборки и критическим уровнем дисбаланса ($N_B = 8000$, $ir = 0{,}0075$, $tr = 0{,}05$). В данных условиях классификатор LightGBM без предварительной аугментации находился

в вырожденном состоянии, демонстрируя F1-macro = 0,03 (фактическое игнорирование минорных классов). Применение CopulaGAN позволило восстановить функциональность модели, подняв значение целевой метрики до 0,81 (абсолютный прирост 0,78). Данный результат доказывает, что при достижении определенного порога объема данных, нейросетевая аугментация способна полностью компенсировать структурные недостатки классификатора, переводя его из состояния неработоспособности в режим эффективного обнаружения атак.

4. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования эффективности генеративно-состязательных сетей для аугментации IoT-трафика сделаны следующие выводы:

1. Оптимальные архитектуры. Установлено, что для табличных данных формата NetFlow наиболее эффективной является модель CopulaGAN, превзошедшая другие нейросетевые подходы (TVAE, CTGAN). При этом выявлено, что классические методы (SMOTE, ROS) сохраняют преимущество на сверхмалых выборках из-за меньшей требовательности к объему обучающих данных.

2. Границы применимости. Экспериментально определен «порог целесообразности» аугментации: она критически важна при высоком дисбалансе ($ir \leq 0{,}015$) и умеренной интенсивности синтеза ($tr \leq 0{,}25$). Показано, что увеличение объема исходных данных само по себе не решает проблему дисбаланса, но создает необходимую базу для качественного обучения GAN-генератора.

3. Восстановление работоспособности моделей. Выявлен значительный синергетический эффект при использовании связки CopulaGAN + LightGBM. В условиях экстремального дефицита данных аугментация позволила восстановить функциональность классификатора, подняв показатель F1-macro с вырожденного уровня 0,03 до эксплуатационного значения 0,81.

4. Практическая значимость. Предложенная методика нейросетевой аугментации позволяет эффективно обучать системы обнаружения вторжений (IDS) в условиях невозможности сбора реального вредоносного трафика, обеспечивая надежную классификацию за счет переноса «бремени дисбаланса» с модели классификатора на этап подготовки данных.

КОНФЛИКТ ИНТЕРЕСОВ / CONFLICT OF INTERESTS

Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflict of interests.

СПИСОК ИСТОЧНИКОВ

1. State of IoT 2025: Number of connected IoT devices growing 14% to 21.1 billion globally. URL: https://iot-analytics.com/number-connected-iot-devices/ (дата обращения: 01.02.2026).
2. IoT Hacking Statistics 2025: The Definitive Report on Threats, Risks & Regulations. URL: https://deepstrike.io/blog/iot-hacking-statistics (дата обращения: 01.02.2026).
3. Nozomi Networks Labs. OT/IoT Cybersecurity Trends & Insights 2025. URL: https://www.nozominetworks.com/ot-iot-cybersecurity-trends-insights-february-2025 (дата обращения: 01.02.2026).
4. Canadian Institute for Cybersecurity. CIC IoT dataset 2023. URL: https://www.unb.ca/cic/datasets/iotdataset-2023.html (дата обращения: 01.02.2026).
5. Kaggle. UNB CIC IOT 2023 Dataset. URL: https://www.kaggle.com/datasets/madhavmalhotra/unb-cic-iot-dataset (дата обращения: 01.02.2026).
6. imbalanced-learn. SMOTE. URL: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html (дата обращения: 01.02.2026).
7. **Xu L., Skoularidou M., Cuesta-Infante A., et al**. Modeling tabular data using conditional

GAN // Advances in neural information processing systems. 2019. Vol. 32. arXiv: 1907.00503.
8. **Zhao Z., Kunar A., Birke R., Chen L. Y.** CTAB-GAN+: Enhancing tabular data synthesis // Frontiers in Big Data. 2024. Vol. 6. P. 1296508.
9. **Camino R., Hammerschmidt C., State R.** Generating multi-categorical samples with generative adversarial networks // arXiv preprint arXiv: 1807.01202. 2018.
10. **Hongwei Ding, Yu Sun, Nana Huang, Xiaohui Cui.** TMG-GAN: Generative adversarial networks-based imbalanced learning for network intrusion detection // IEEE Transactions on Information Forensics and Security. 2023. Vol. 19. P. 1156–1167. DOI: 10.1109/TIFS.2023.3331240
11. Synthetic Data Vault. CopulaGANSynthesizer. URL: https://docs.sdv.dev/sdv/single-table-data/modeling/synthesizers/copulagansynthesizer (дата обращения: 01.02.2026).
12. Synthetic Data Vault. CTGANSynthesizer. URL: https://docs.sdv.dev/sdv/single-table-data/modeling/synthesizers/ctgansynthesizer (дата обращения: 01.02.2026).
13. Synthetic Data Vault. TVAESynthesizer. URL: https://docs.sdv.dev/sdv/single-table-data/modeling/synthesizers/tvaesynthesizer (дата обращения: 01.02.2026).
14. LightGBM's documentation. URL: https://lightgbm.readthedocs.io/en/stable/ (дата обращения: 01.02.2026).
15. scikit-learn. RandomForestClassifier. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (дата обращения: 01.02.2026).
16. XGBoost Documentation. URL: https://xgboost.readthedocs.io/en/stable/ (дата обращения: 01.02.2026).

REFERENCES

1. State of IoT 2025: Number of connected IoT devices growing 14% to 21.1 billion globally. URL: https://iot-analytics.com/number-connected-iot-devices/ (accessed: 01.02.2026).
2. IoT Hacking Statistics 2025: The Definitive Report on Threats, Risks & Regulations. URL: https://deepstrike.io/blog/iot-hacking-statistics (accessed: 01.02.2026).
3. Nozomi Networks Labs. OT/IoT Cybersecurity Trends & Insights 2025. URL: https://www.nozominetworks.com/ot-iot-cybersecurity-trends-insights-february-2025 (accessed: 01.02.2026).
4. Canadian Institute for Cybersecurity. CIC IoT dataset 2023. URL: https://www.unb.ca/cic/datasets/iotdataset-2023.html (accessed: 01.02.2026).
5. Kaggle. UNB CIC IOT 2023 Dataset. URL: https://www.kaggle.com/datasets/madhavmalhotra/unb-cic-iot-dataset (accessed: 01.02.2026).
6. imbalanced-learn. SMOTE. URL: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html (accessed: 01.02.2026).
7. **Xu L., Skoularidou M., Cuesta-Infante A., et al**. Modeling tabular data using conditional GAN. *Advances in neural information processing systems*. 2019. Vol. 32. arXiv: 1907.00503.
8. **Zhao Z., Kunar A., Birke R., Chen L. Y.** CTAB-GAN+: Enhancing tabular data synthesis. *Frontiers in Big Data*. 2024. Vol. 6, pp. 1296508.
9. **Camino R., Hammerschmidt C., State R.** Generating multi-categorical samples with generative adversarial networks. *arXiv preprint arXiv: 1807.01202*. 2018.
10. **Hongwei Ding, Yu Sun, Nana Huang, Xiaohui Cui.** TMG-GAN: Generative adversarial networks-based imbalanced learning for network intrusion detection. *IEEE Transactions on Information Forensics and Security*. 2023. Vol. 19, pp. 1156–1167. DOI: 10.1109/TIFS.2023.3331240
11. Synthetic Data Vault. CopulaGANSynthesizer. URL: https://docs.sdv.dev/sdv/single-table-data/modeling/synthesizers/copulagansynthesizer (accessed: 01.02.2026).
12. Synthetic Data Vault. CTGANSynthesizer. URL: https://docs.sdv.dev/sdv/single-table-data/modeling/synthesizers/ctgansynthesizer (accessed: 01.02.2026).
13. Synthetic Data Vault. TVAESynthesizer. URL: https://docs.sdv.dev/sdv/single-table-data/modeling/synthesizers/tvaesynthesizer (accessed: 01.02.2026).

14. LightGBM's documentation. URL: https://lightgbm.readthedocs.io/en/stable/ (accessed: 01.02.2026).
15. scikit-learn. RandomForestClassifier. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (accessed: 01.02.2026).
16. XGBoost Documentation. URL: https://xgboost.readthedocs.io/en/stable/ (accessed: 01.02.2026).

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

ПЛАТОНОВ Владимир Владимирович – канд. техн. наук, доцент, Санкт-Петербургский политехнический университет Петра Великого, Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29
E-mail: plato@ibks.spbstu.ru
ORCID: 0000-0002-9899-2778

PLATONOV Vladimir V. – Candidate of Engineering Sciences, Associate Professor, Peter the Great St. Petersburg Polytechnic University, Russia, 195251, St. Petersburg, Polytechnicheskaya str., 29

СКИБА Дарослав Александрович – программист, АО «ИнфоТеКС», Россия, 125167, Москва, ул. Викторенко, д. 9, стр. 1
E-mail: daroslav.skiba@yandex.ru
ORCID: 0009-0004-9032-4961

SKIBA Daroslav A. – Programmer, JSC "InfoTeX", Russia, 125167, Moscow, Viktorenko str., 9, build.1