

УДК 004.8

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОБУЧЕНИЯ БЕЗ УЧИТЕЛЯ ДЛЯ ДЕТЕКЦИИ АНОМАЛИЙ В МНОГОМЕРНОЙ ПРОМЫШЛЕННОЙ ТЕЛЕМЕТРИИ

**Морозова Вероника Сергеевна,**

магистрант отделения интеллектуальных кибернетических систем Обнинский институт атомной энергетики - филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ»  
morozovavs25@oiate.ru

**Потапов Андрей Евгеньевич,**

кандидат физико-математических наук, доцент кафедры ИУК5 Калужский филиал федерального государственного автономного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»  
potapovae@bmstu.ru

### Аннотация

В настоящее время растет распространение данных в области промышленной телеметрии, и соответственно, для улучшения стратегий поиска аномалий в производственных циклах необходимо пересматривать и оптимизировать стратегии их поиска. В машинном обучении существует множество алгоритмов для поиска аномалий во временных рядах. С помощью методов машинного обучения можно обнаружить различные закономерности неисправности оборудования и затем предотвратить выход из строя машины. Однако отдельные алгоритмы могут давать ложные результаты в различных наборах данных, и ни один алгоритм не может работать оптимально во всех сценариях. В рамках этого исследования представлен сравнительный анализ алгоритмов машинного обучения без учителя для поиска аномалий в производственных циклах. В исследовании используются плотностные, геометрические методы, ансамбли машинного обучения.

**Ключевые слова:** обнаружение аномалий, методы машинного обучения без учителя, предиктивное обслуживание.

## COMPARATIVE ANALYSIS OF UNSUPERVISED LEARNING METHODS FOR ANOMALY DETECTION IN MULTIDIMENSIONAL INDUSTRIAL TELEMETRY

**Morozova Veronika Sergeevna,**

Master's Student at the Department of Intelligent Cybernetic Systems, Obninsk Institute of Atomic Energy, a branch of the Federal State Autonomous Educational Institution of Higher Education National Research Nuclear University MEPhI

**Potapov Andrey Evgenievich,**

Candidate of Physico-Mathematical Sciences,

Associate Professor of the Department of IUK5 Kaluga Branch of the Federal State

Autonomous Educational Institution of Higher Education «Bauman Moscow State Technical University (National Research University)»

---

## ABSTRACT

---

Currently, the proliferation of data in the field of industrial telemetry is increasing, and accordingly, in order to improve anomaly detection strategies in production cycles, they need to be reviewed and optimized. In machine learning, there are many algorithms for finding anomalies in time series. Using machine learning methods, it is possible to detect various patterns of equipment malfunction and then prevent machine failure. However, individual algorithms can produce false results in different datasets, and no single algorithm can perform optimally in all scenarios. This study presents a comparative analysis of unsupervised machine learning algorithms to search for anomalies in production cycles. The research uses density, geometric methods, and machine learning ensembles.

---

**Keywords:** anomaly detection, unsupervised machine learning methods, predictive maintenance

---

### Введение

В этом исследовании проводится сравнительный анализ методов машинного обучения без учителя с реальными данными, которые получены с датчиков промышленного оборудования, затем они загружаются в базу данных. Характеристики датчиков – это различные физические параметры, например, токи L1, L2, L3, напряжения L1, L2, L3, мгновенная и реактивная мощности, грузовой момент и другие.

Актуальность заключается в том, что современное промышленное оборудование генерирует огромные потоки данных, ручной анализ которых невозможен. Своевременное обнаружение аномалий и скрытых паттернов в работе узлов позволяет предотвратить дорогостоящие поломки, сократить время внеплановых простоев и оптимизировать энергопотребление. Это обуславливает необходимость применения ML подходов, способных адаптироваться к изменяющимся условиям оборудования без предварительной разметки данных.

Использование методов машинного обучения без учителя особенно важно, так как в реальных производственных условиях данные чаще всего не размечены: заранее неизвестно рабочий или аномальный цикл. Обнаружение аномалий включает в себя процесс распознавания закономерностей в данных, которые отличаются от ожидаемого поведения [1].

В работе решается задача автоматического выделения стабильных режимов работы и идентификации нетипичных циклов, что служит основой для создания систем предиктивного обслуживания. Проблема обнаружения аномалий, обсуждаемая в этом исследовании, нельзя решить с помощью методов машинного обучения с учителем, так как основной целью является выявление аномалий без каких-либо предварительных знаний о том, что является нормальным рабочим циклом или аномальным.

Также в работе применен динамический подход к настройке гиперпараметров для каждой фазы цикла отдельно.

Материалы и методы исследования

## Обзор алгоритмов

Необходимо отметить, что циклы производственного оборудования представляют собой временные ряды. Каждый цикл состоит, в свою очередь, из нескольких фаз. Для поиска аномалий циклов необходимо проверять каждую фазу на наличие аномалии в алгоритме, соответственно, при наличии отклоняющегося поведения у фазы цикл также помечается как аномальный.

Одни из методов, которые применяются в исследовании, — это методы кластеризации.

Кластеризация – это процесс исследования множества «точек» и их группировки в «кластеры» согласно некоторой метрике. Идея в том, что точки, попавшие в один кластер, находятся недалеко друг от друга, а точки из разных кластеров разделены большим расстоянием [2].

В данном разделе осуществляется обзор алгоритмов для поиска аномалий: HDBSCAN, иерархическая кластеризация (agglomerative clustering), One class SVM, Isolation forest, Angle-based outlier detection (ABOD), Local outlier factors (LOF).

### 1. HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise)

Данный алгоритм целесообразен для обнаружения кластеров различной формы и размера без необходимости жесткого задания радиуса соседства, что важно для анализа различных фаз производственного цикла. Алгоритм расширяет возможности DBSCAN путем преобразования его в метод иерархической кластеризации, а затем использует метод выделения плоской кластеризации, основанный на стабильности кластеров. Алгоритм HDBSCAN позволяет идентифицировать и пометить определенные точки данных как шумовые, что позволяет избежать необходимости относить эти выборки к определенному кластеру. Большинство из этих зашумленных выборок, как правило, расположены в непосредственной близости от границы между двумя различными группами, или они могут быть расположены достаточно далеко от любого идентифицированного кластера.

### 2. Иерархическая кластеризация

Иерархические, или агломеративные алгоритмы сначала помещают каждую точку в отдельный кластер. Затем «близкие» кластеры объединяются с использованием одного из многих определений «близости». Процесс объединения прекращается, когда дальнейшее объединение приводит к кластерам, которыми по каким-то причинам нежелательны [2].

### 3. One class SVM

В данной работе используется One class SVM, так как достаточно разделить датасет, содержащий циклы с аномальным поведением и циклы нормальной работой.

Одна из реализаций SVM одного класса заключается в отображении входных данных в многомерное пространство объектов, а затем в размещении всех или большей части данных в гиперсфере.

Идея этого метода заключается в том, что объем гиперсферы сводится к минимуму, и все выборки данных, которые не попадают в гиперсферу, рассматриваются как аномалии [3, 4, 5].

### 4. Isolation forest

Алгоритм Isolation Forest, широко применяемый метод обнаружения аномалий, использует неконтролируемый подход, используя принцип случайного рекурсивного разделения пространства признаков [6, 7]. Он обеспечивает линейную временную сложность и требует минимального объема памяти даже при применении к большим наборам данных и с большим количеством нерелевантных объектов. Данный алгоритм изолирует аномалии путем построения деревьев решений, называемых изолирующими деревьями, которые постепенно разделяют точки данных с помощью случайных разбиений.

Аномалии определяются как редкие наблюдения, которые легко выделяются из большинства точек данных [8].

### 5. Angle-based outlier detection (ABOD)

Этот метод основан на идее отслеживания угла, образованного набором из любых трех точек данных в пространстве многовариантных объектов. Разница в величине угловой границы оказывается различной для точек с выбросами и нормальных точек. Обычно наблюдаемая дисперсия выше для более низких точек, чем для выбросов, следовательно, такая мера помогает нам по-разному группировать нормальные и более высокие точки. Данный метод довольно хорошо работает в многомерном пространстве, в отличие от других методов измерения расстояний, которые страдают от “проклятия размерности”. Расстояние между любыми двумя точками в пространстве высокой размерности почти одинаково. В таких сценариях углы могут дать лучшее представление о близости [9, 10].

### 6. Local outlier factors (LOF)

Для многих сценариев более целесообразно присваивать каждому объекту степень выброса. Эта степень называется коэффициентом локального выброса (LOF) объекта. Она является локальной в том смысле, что степень зависит от того, насколько изолирован объект по отношению к окружающей среде [11].

#### 2. Экспериментальная часть

##### 2.1. Набор данных телеметрии

В качестве набора данных используется датасет, в котором собраны результаты работы датчиков промышленных машин (например, токи L1, L2, L3, мгновенное значение нагрузки вспомогательного подъема замыкания и главного подъема, потребление электроэнергии, грузовой момент, активная и реактивная мощности, скорость привода и другие параметры).

#### 2. Библиотеки

Для обработки используется библиотеки Python scikit-learn (HDBSCAN, AgglomerativeClustering, LocalOutlierFactor, OneClassSVM, IsolationForest), pyod (ABOD).

#### 3. Метрики

Для проверки моделей используются следующие метрики: количество полученных кластеров (для методов кластеризации), шум, коэффициент силуэта кластера, latency.

#### 4. Результаты работы алгоритмов

Динамическая настройка гиперпараметров осуществляется в зависимости от фазы цикла.

#### HDBSCAN

Для фаз 2 и 4 использовались жесткие критерии плотности (cluster\_selection\_epsilon=0.5) и метод eom, что позволило выделить устойчивые режимы работы оборудования.

Для фаз 1 и 3: cluster\_selection\_epsilon увеличен до 1.8–2.0, а метод leaf позволил идентифицировать даже небольшие локальные сгущения данных.

store\_centers - medoid: необходимо для нахождения реального исторического цикла, наиболее репрезентативно описывающий состояние системы в данном кластере.

Таблица 1 – Результаты HDBSCAN

Фаза цикла	Количество кластеров	Шум	Силуэтный коэффициент	Latency (мс)
1	3	0.096	0.232	1.25
2	2	0.0	0.802	1.10
3	2	0.096	0.332	1.15

4	2	0.0	0.792	1.12
---	---	-----	-------	------

### Agglomerative clustering

Здесь настроены следующие гиперпараметры:

Количество кластеров (`n_clusters`): для фаз 2 и 4 задано значение 2, что соответствует двум основным направлениям процесса. Для фаз 1 и 3 количество кластеров было увеличено до 3 с целью изоляции экстремальных выбросов в отдельные группы и предотвращения их влияния на усредненные характеристики основных режимов.

Метрика расстояния (`metric: 'manhattan'`): использование манхэттенского расстояния (L1-норма) обусловлено его устойчивостью к единичным выбросам в показаниях датчиков.

Метод связи (`linkage: 'complete'`): это обеспечило формирование компактных и четко разделенных групп.

Таблица 2 - Результаты Agglomerative clustering

Фаза цикла	Количество кластеров	Силуэтный коэффициент	Latency (мс)
1	3	0.213	2.85
2	2	0.802	2.50
3	3	0.239	2.60
4	2	0.781	2.55

### One class SVM

Настроены следующие параметры:

`kernel: 'rbf'` - выбор радиально-базисной функции (гауссово ядро) обусловлен сложной, нелинейной структурой распределения данных телеметрии.

Параметр `nu`: для фаз 2 и 4 установлено значение 0.02, так как процессы в эти периоды детерминированы. Для фаз 1 и 3 значение увеличено до 0.05, что позволяет алгоритму быть более адаптивным к естественной вариативности пусковых и стопорных режимов.

Таблица 3 - Результаты One class SVM

Фаза цикла	Шум	Силуэтный коэффициент	Latency (мс)
1	0.173	0.09	3.40
2	0.059	0.034	3.10
3	0.192	0.081	3.25
4	0.019	-0.197	3.15

### Isolation forest

Здесь были применены параметры по умолчанию в соответствии с документацией.

Таблица 4 - Результат Isolation forest

Фаза цикла	Шум	Силуэтный коэффициент	Latency (мс)
1	0.058	0.439	195.5
2	0.039	0.25	192.0

3	0.058	0.429	194.0
4	0.038	0.084	193.5

### ABOD

n\_neighbors: для фаз 1 и 3 значение равно 5, что позволяет алгоритму реагировать на локальные угловые отклонения. Для фаз 2 и 4 количество соседей увеличено до 10 для обеспечения более стабильной оценки направления векторов в условиях линейных зависимостей параметров.

method: 'fast': использование оптимизированного метода fast позволило ограничить расчет угловых коэффициентов только ближайшим окружением точки.

metric: 'minkowski', p: 2: применение стандартной евклидовой метрики в рамках расчета векторов признаков обеспечило консистентность результатов с другими методами и позволило корректно оценивать взаимное расположение циклов в пространстве после процедуры стандартизации.

Таблица 5 - Результаты ABOD

Фаза цикла	Шум	Силуэтный коэффициент	Latency (мс)
1	0.058	0.343	45.2
2	0.059	0.08	42.1
3	0.058	0.402	44.5
4	0.058	0.102	43.0

### LOF

n\_neighbors: для фаз 1 и 3 выбрано значение 5, что обеспечивает высокую чувствительность к локальным отклонениям в разреженных облаках данных. Для фаз 2 и 4, характеризующихся высокой плотностью и линейной структурой, количество соседей увеличено до 10, чтобы алгоритм мог корректно оценивать среднюю плотность вдоль рабочих траекторий.

contamination 0.05: настроен для поиска наиболее выраженных "пустот" в распределении плотности, которые соответствуют нетипичным циклам работы. На рис. 1 представлено разделение циклов с помощью LOF.

Таблица 6 - Результаты LOF

Фаза	Шум	Силуэтный коэффициент	Latency (мс)
1	0.058	0.327	4.10
2	0.059	0.203	3.85
3	0.058	0.382	3.90
4	0.058	0.202	3.88

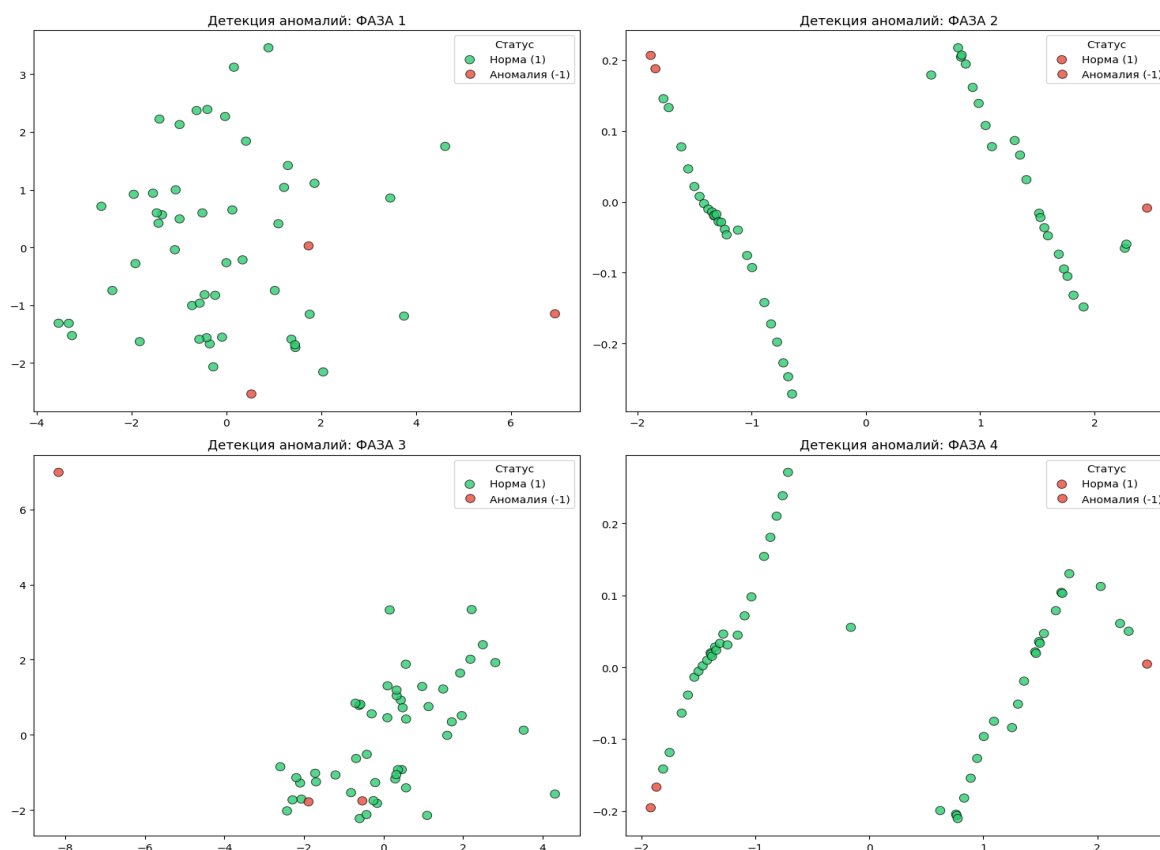


Рисунок 1 – Результаты выделения аномалий LOF

Таким образом, один из приемлемых результатов показал алгоритм HDBSCAN. Благодаря способности работать с переменной плотностью, он обеспечил стабильное разделение на кластеры в фазах со стабильной нагрузкой (коэффициент силуэта 0.802). При этом метод продемонстрировал минимальную вычислительную трудоемкость. Также стоит отметить эффективность метода LOF. Isolation Forest потребовал больших затрат по времени, также One Class SVM не отработал эффективно.

#### Заключение

В этой работе проведен сравнительный анализ алгоритмов машинного обучения без учителя для обнаружения аномалий в многомерных временных данных промышленной телеметрии. В исследовании используются методы HDBSCAN, Agglomerative clustering, One class SVM, Isolation forest, ABOD, LOF для обработки реальных данных и сравниваются их результаты работы. Результаты демонстрируют, что для автоматического выделения стабильных технологических состояний наиболее эффективным по результатам является HDBSCAN, LOF показал себя как оптимальное решение для систем реального времени, сочетая высокую скорость обработки с качественной локальной детекцией, ABOD также показал приемлемые результаты, но также чувствителен к параметру  $n\_neighbors$ . Isolation Forest, несмотря на высокую точность в зашумленности данных, оказался затратным по времени. One class SVM некорректно отработал, так как получен отрицательный коэффициент. Поскольку ни один из алгоритмов не является универсальным для всех фаз цикла, дальнейшее развитие систем предиктивного обслуживания должно базироваться на ансамблях алгоритмов, включающих не только методы машинного обучения.

#### Список литературы:

1. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection : A Survey // ACM Computing Survey. - 2009. - №41 (3) С. 1-58. DOI: 10.1145/1541880.1541882

2. Лесковец, Ю. Анализ больших наборов данных / Ю. Лесковец, А. Раджараман, Дж. Д. Ульман; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2016. — 498 с
3. Sutharshan Rajasegarar; Christopher Leckie; James C. Bezdek; Marimuthu Palaniswami Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks // IEEE Transactions on Information Forensics and Security. - 2010. - №5 (3). - С. 518 - 533. DOI:10.1109/TIFS.2010.2051543
4. D. Wang, D. S. Yeung and E. C. C. Tsang Structured one-class classification // IEEE Transactions on Systems, Man and Cybernetics. - 2006. - №36 (6). - С. 1283 - 1295. DOI: 10.1109/tsmcb.2006.876189
5. D. M. J. Tax and R. P. W. Duin Support vector data description // Machine Learning. - 2004. - №54 (1). - С. 45-66. DOI: 10.1023/B:MACH.0000008084.60811.49
6. Liu, F.T.; Ting, K.M.; Zhou, Z.H Isolation forest // Eighth IEEE International Conference on Data Mining. - 2008. - С. 413-422. DOI: 10.1109/ICDM.2008.17
7. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation-Based Anomaly Detection // ACM Transactions on Knowledge Discovery from Data. - 2012. - №6 (3). - С. 1-39. DOI: 10.1145/2133360.2133363
8. M. S. Binetti, V. F. Uricchio, C. Massarelli Isolation Forest for Environmental Monitoring: A Data-Driven Approach to Land Management // Environments. - 2025. - №12 (4). DOI: 10.3390/environments12040116
9. Detecting Outliers with Angle-based Techniques in Python // Digital Ocean URL: <https://blog.paperspace.com/outlier-detection-with-abod/> (дата обращения: 10.04.2026).
10. Kriegel H.-P., Schubert M., Zimek A. Angle-Based Outlier Detection in High-Dimensional Data // Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08). - New York, NY, USA : ACM, 2008. - С. 444-452. DOI: 10.1145/1401890.1401946
11. Breunig, M. M. LOF: Identifying Density-Based Local Outliers / M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander // Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD '00). - 2000. - Т. 29, № 2. - С. 93-104. DOI:10.1145/342009.335388.

**References:**

1. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection : A Survey // ACM Computing Survey. - 2009. - Vol. 41 (3) - pp. 1-58. DOI: 10.1145/1541880.1541882
2. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). Mining of massive datasets (2nd ed.). Cambridge University Press.
3. S. Rajasegarar; C. Leckie; J. C. Bezdek; M. Palaniswami Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks // IEEE Transactions on Information Forensics and Security. - 2010. - Vol. 5 (3). - pp. 518 - 533. DOI:10.1109/TIFS.2010.2051543

4. D. Wang, D. S. Yeung and E. C. C. Tsang Structured one-class classification // IEEE Transactions on Systems, Man and Cybernetics. - 2006. - Vol. 36 (6). - pp. 1283 - 1295. DOI: 10.1109/tsmcb.2006.876189
5. D. M. J. Tax and R. P. W. Duin Support vector data description // Machine Learning. - 2004. - Vol. 54 (1). - pp. 45-66. DOI: 10.1023/B:MACH.0000008084.60811.49
6. Liu, F.T.; Ting, K.M.; Zhou, Z.H Isolation forest // Eighth IEEE International Conference on Data Mining. - 2008. - C. 413–422. DOI: 10.1109/ICDM.2008.17
7. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation-Based Anomaly Detection // ACM Transactions on Knowledge Discovery from Data. - 2012. - Vol. 6 (3). - pp. 1-39. DOI: 10.1145/2133360.2133363
8. M. S. Binetti, V. F. Uricchio, C. Massarelli Isolation Forest for Environmental Monitoring: A Data-Driven Approach to Land Management // Environments. - 2025. - Vol. 12 (4). DOI: 10.3390/environments12040116
9. Detecting Outliers with Angle-based Techniques in Python // DigitalOcean URL: <https://blog.paperspace.com/outlier-detection-with-abod/> (дата обращения: 10.04.2026).
10. Kriegel H.-P., Schubert M., Zimek A. Angle-Based Outlier Detection in High-Dimensional Data // Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08). - New York, NY, USA : ACM, 2008. - pp. 444–452. DOI: 10.1145/1401890.1401946
11. Breunig, M. M. LOF: Identifying Density-Based Local Outliers / M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander // Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD '00). - 2000. - Vol. 29 (2). - pp. 93-104. DOI:10.1145/342009.335388.