

Выявление аномалий в экспериментальных данных методами машинного обучения

Выполнил:

Перчихин Олег Игоревич

Дипломный руководитель:

Мочалова Юлия Дмитриевна
ФИЦ ИУ РАН, ВМК МГУ

Предметная область и актуальность применения:

Компоновка, размещение элементов и трассировка печатной платы, выполненные без учета требований ЭМС, в большинстве случаев являются причиной сбоев и отказов. В случае выявления помехи возникает проблема определения ее локализации. Для этого могут быть использованы измерительные приборы – сканеры ближнего электромагнитного поля.

Симуляция и моделирование объектов исследования для валидации эксперимента зачастую не представляются возможны или целесообразными, в виду сложности их реализации и значительных ресурсов затрачиваемых при проектировании.

В связи с этим предлагается рассмотреть возможность применения методов машинного обучения для выявления аномалий в экспериментальных данных.

Категориальный аппарат

В данной работе под аномалиями будут пониматься данные, которые не соответствуют общему поведению набора данных и восприниматься как выбросы.



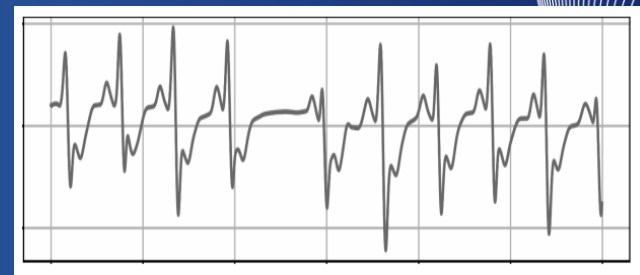
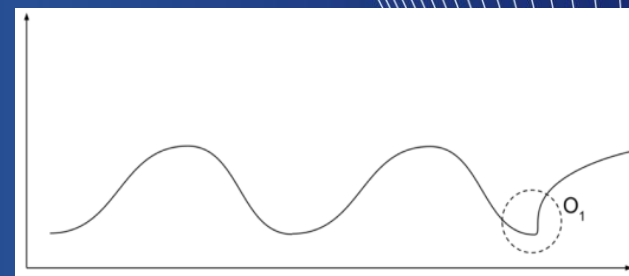
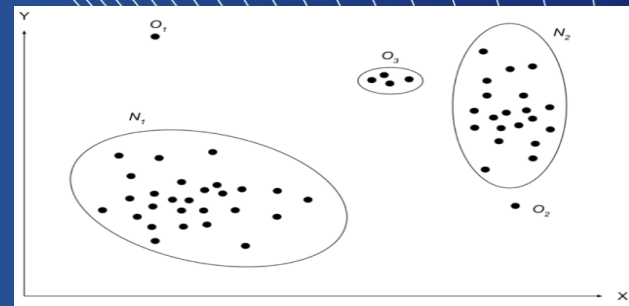
Категориальный аппарат

Выбросы

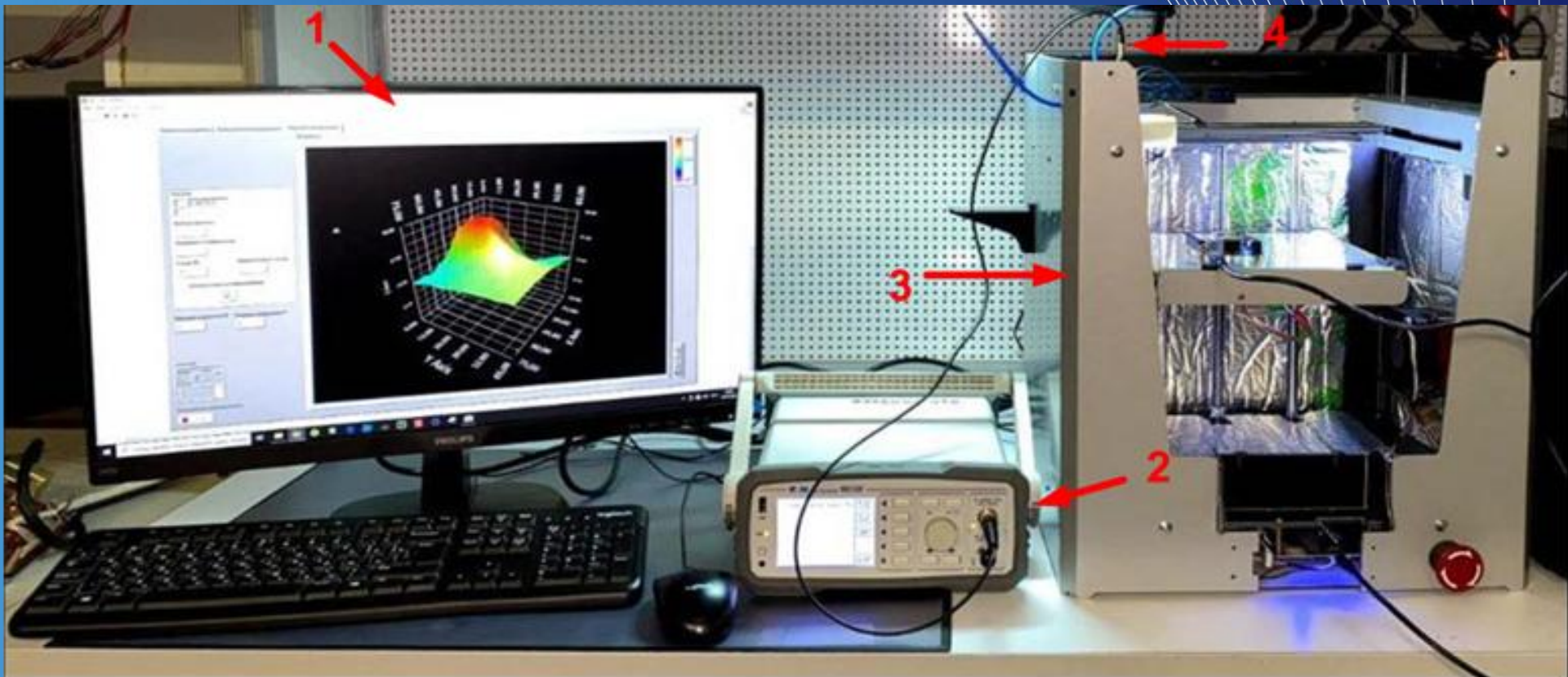
Глобальные

Контекстные

Коллективные



Макет информационно-измерительной системы



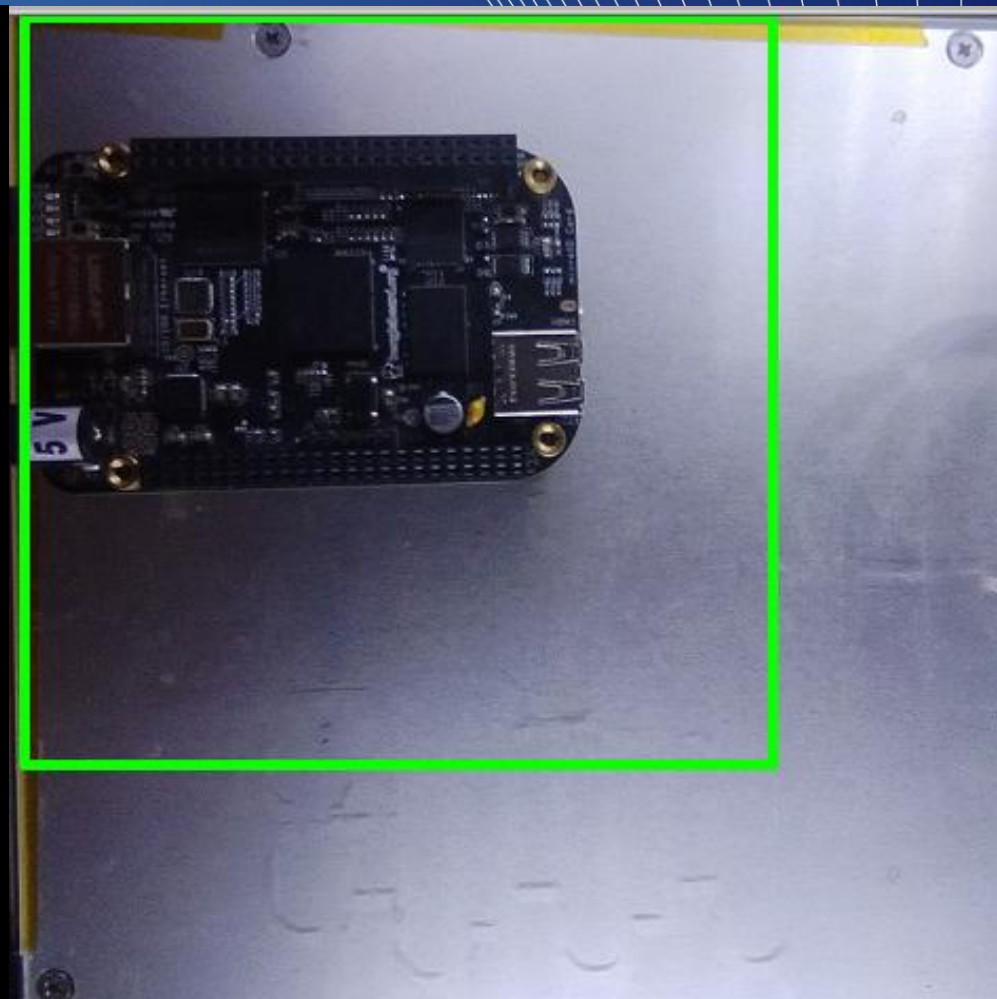
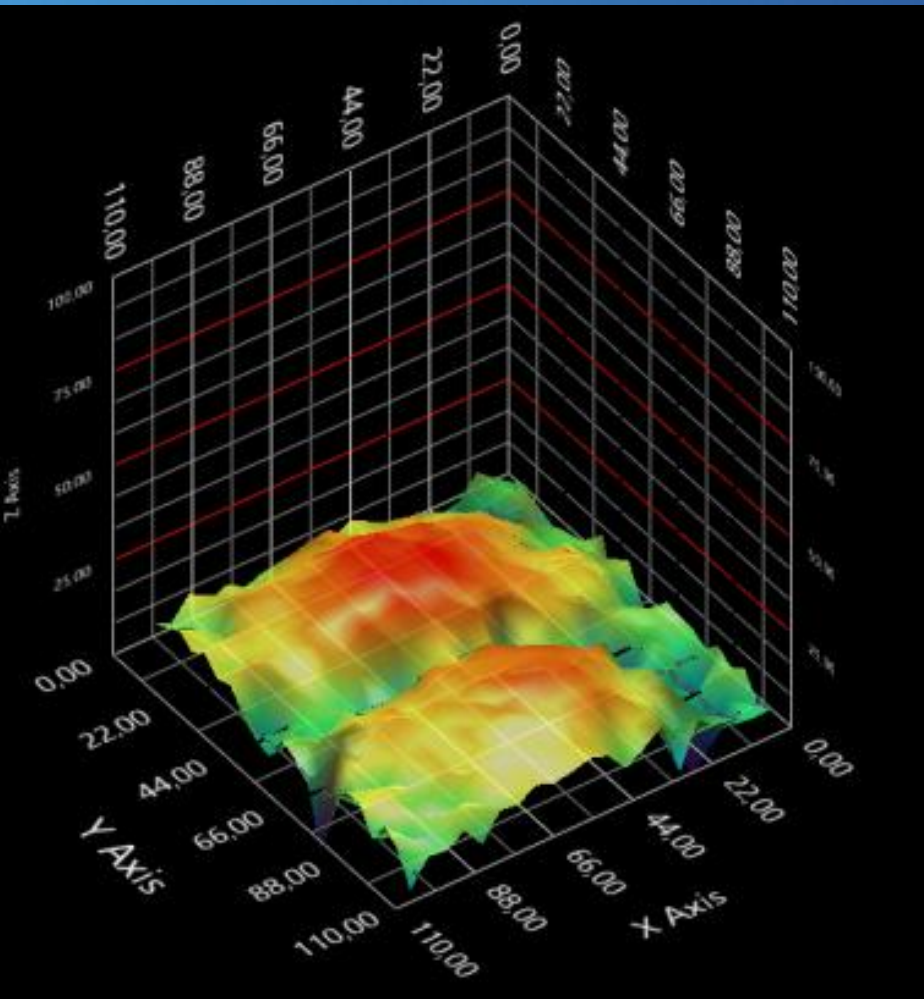
1-ПК с прикладным ПО

2-Приемник ЭМИ

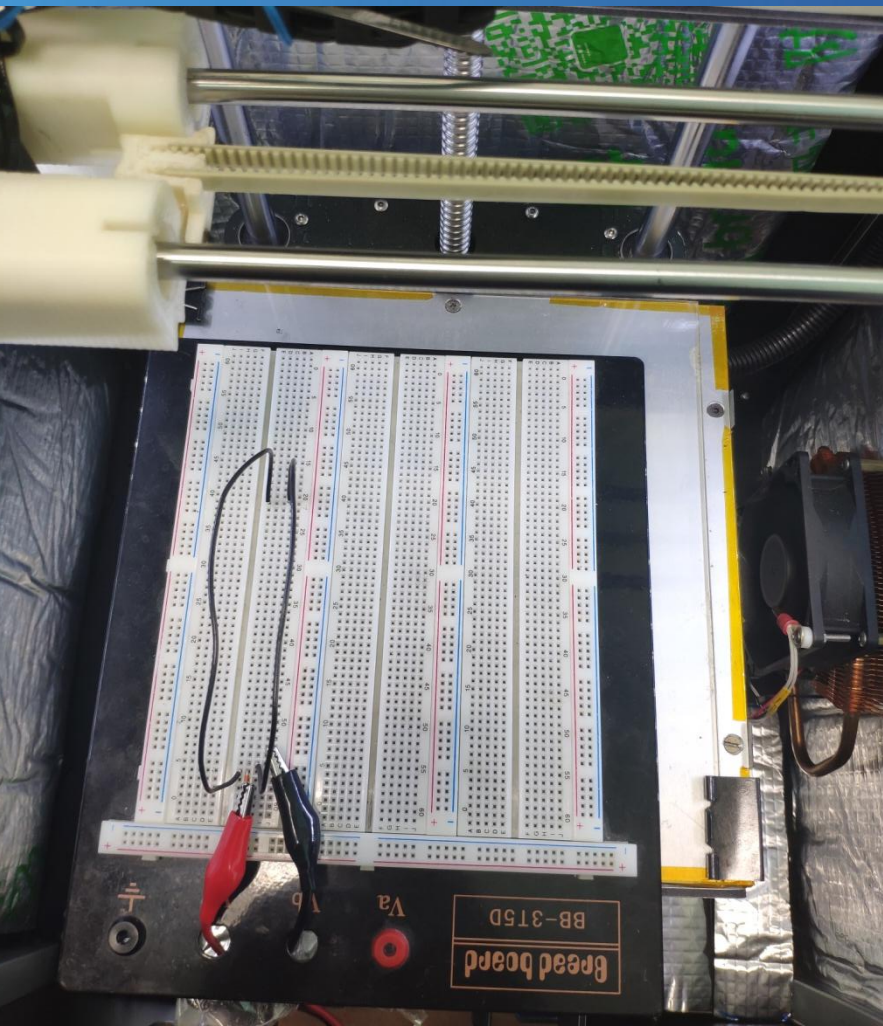
3-Роботизированная системы позиционирования датчика

4-Датчик ближней зоны

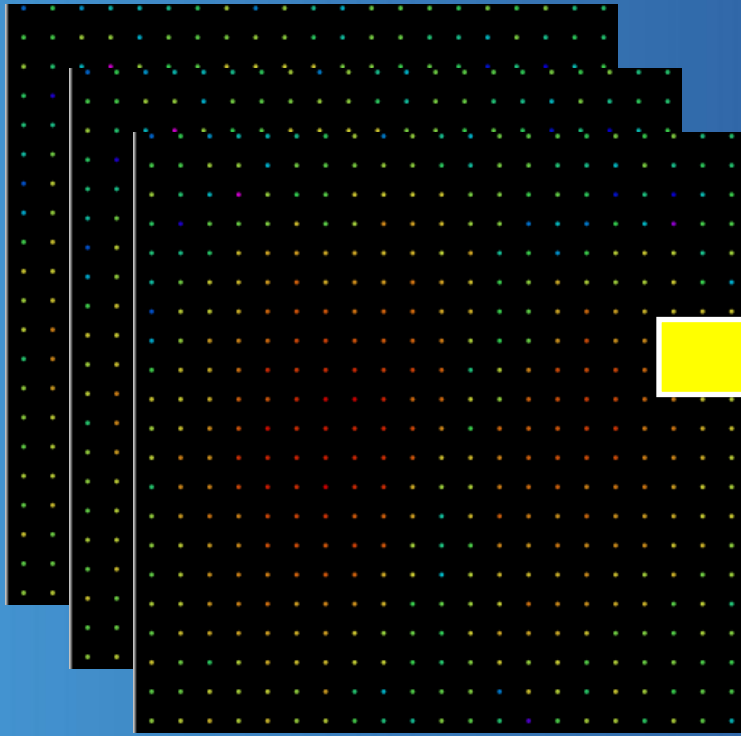
Картирование ЭМП



Тестовый объект исследовательская (DUT)




Протоколирование результатов

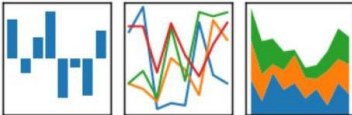


	A	B	C	D	E	F
1		Частота, Гц	дБ/мкВ	Координата X, мм	Координата Y, мм	Зона сканирования
2	Измерение № 1	50000000	30,39	1	1	1
3	Измерение № 2	100000000	31,92	1	1	1
4	Измерение № 3	150000000	33,28	1	1	1
5	Измерение № 4	200000000	29,88	1	1	1
6	Измерение № 5	250000000	32,81	1	1	1
7	Измерение № 6	500000000	33,52	2	1	1
8	Измерение № 7	1000000000	29,45	2	1	1
9	Измерение № 8	1500000000	32,11	2	1	1
10	Измерение № 9	2000000000	31,47	2	1	1

Анализ/
Описательная статистика



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$


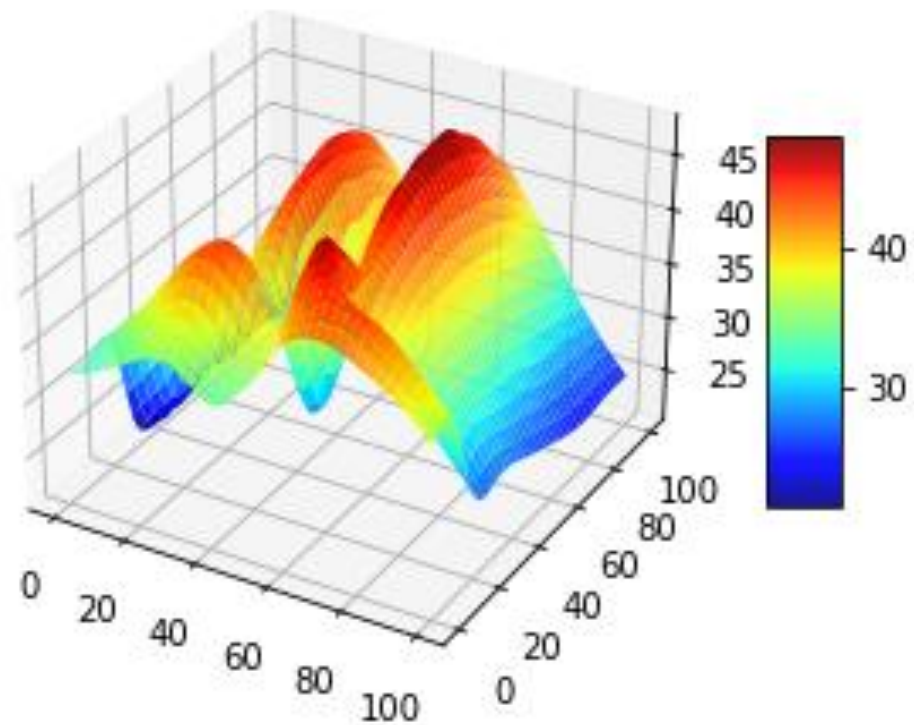
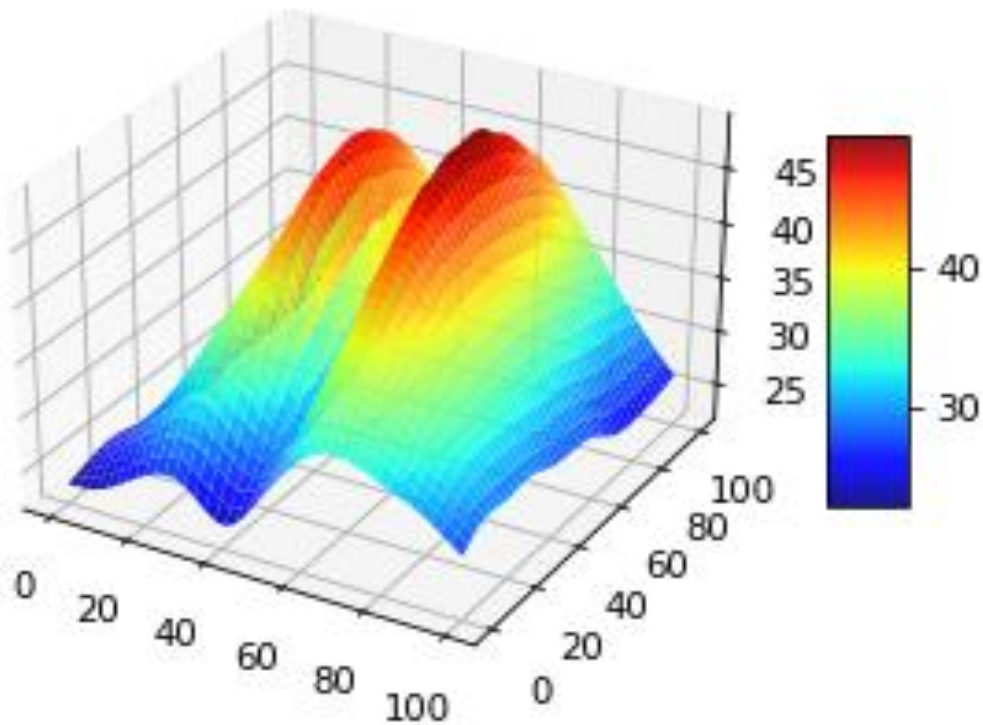
Накопление



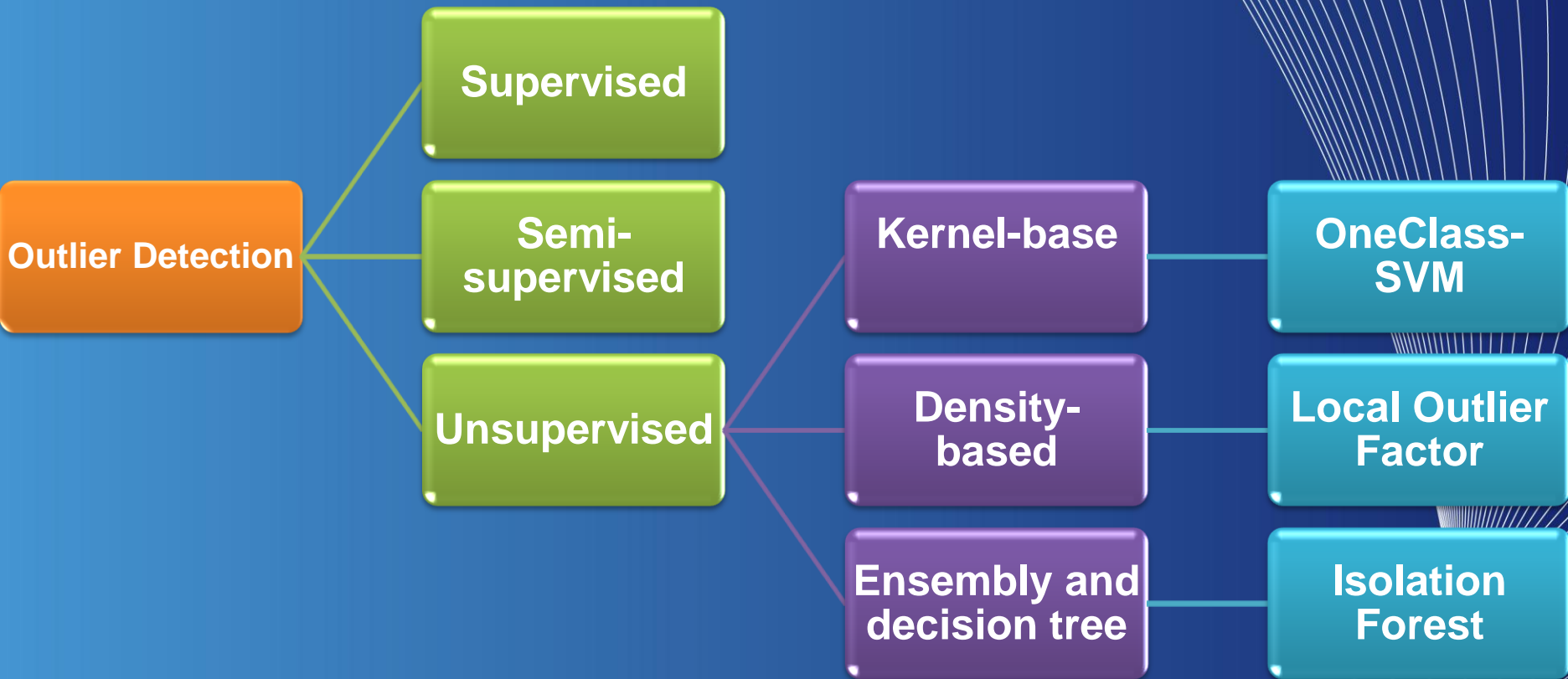
Пример визуализации результатов:

«НОМИНАЛЬНОЕ»

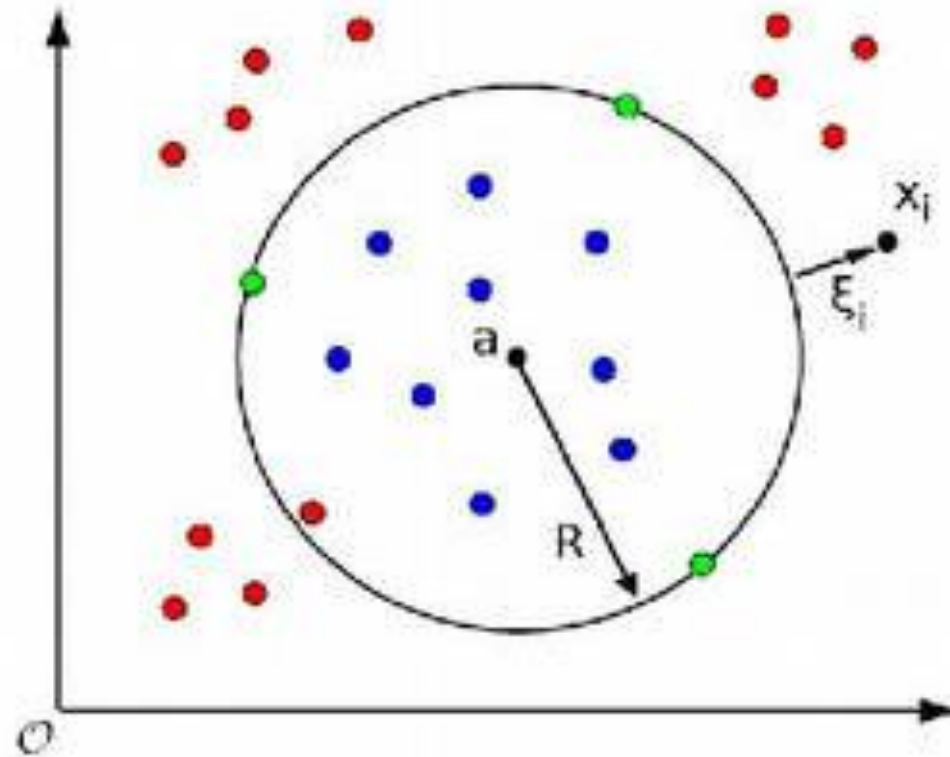
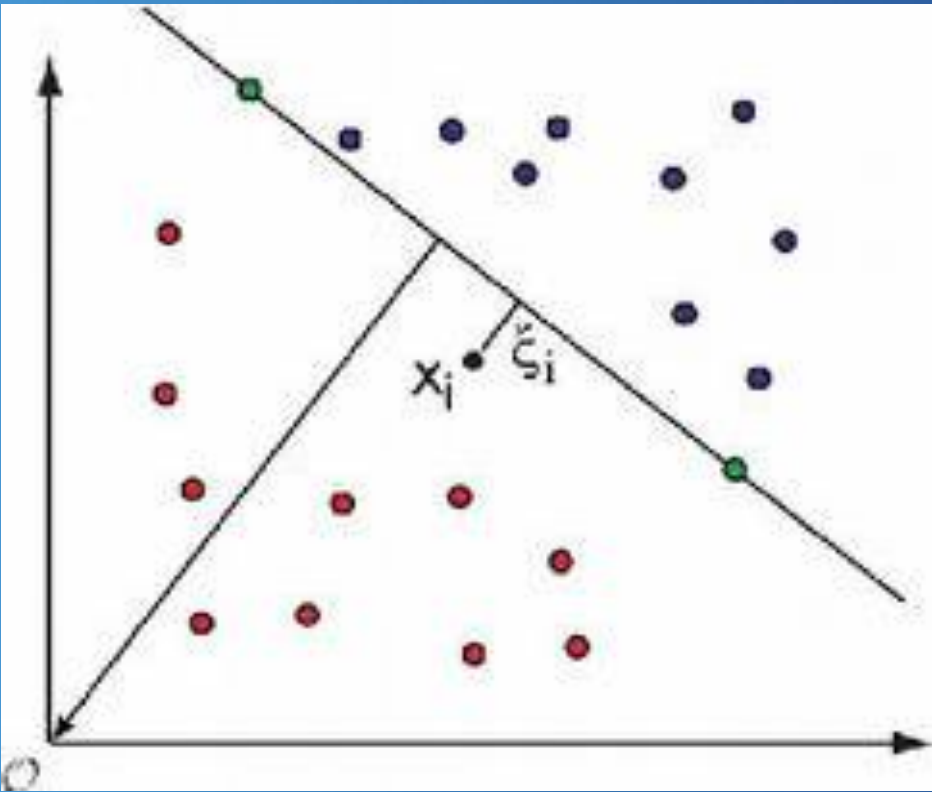
«аномальное»



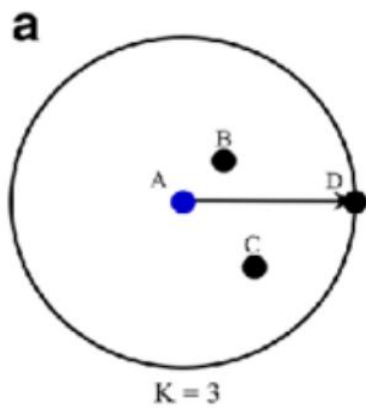
Использованные методы



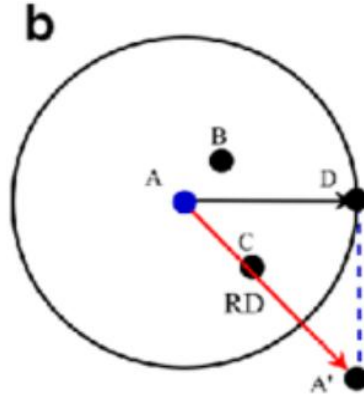
Метод опорных векторов для одного класса One Class-SVM



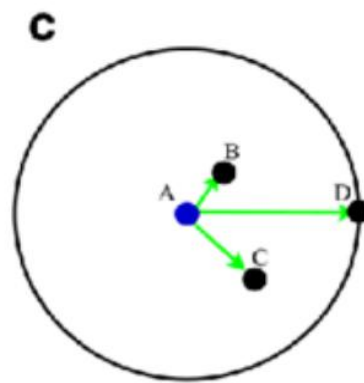
Фактор Локального Выброса Local Outlier Factor



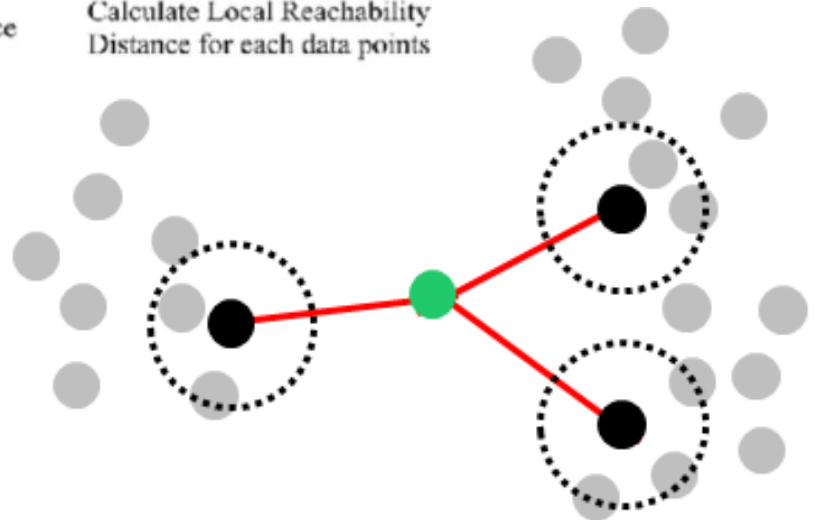
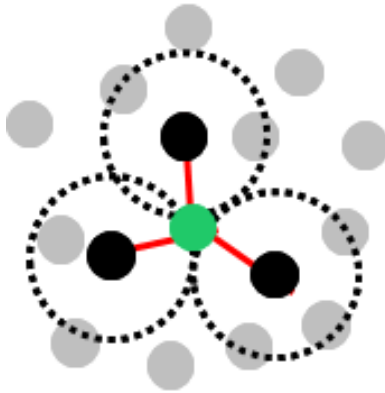
Determine the number of K .



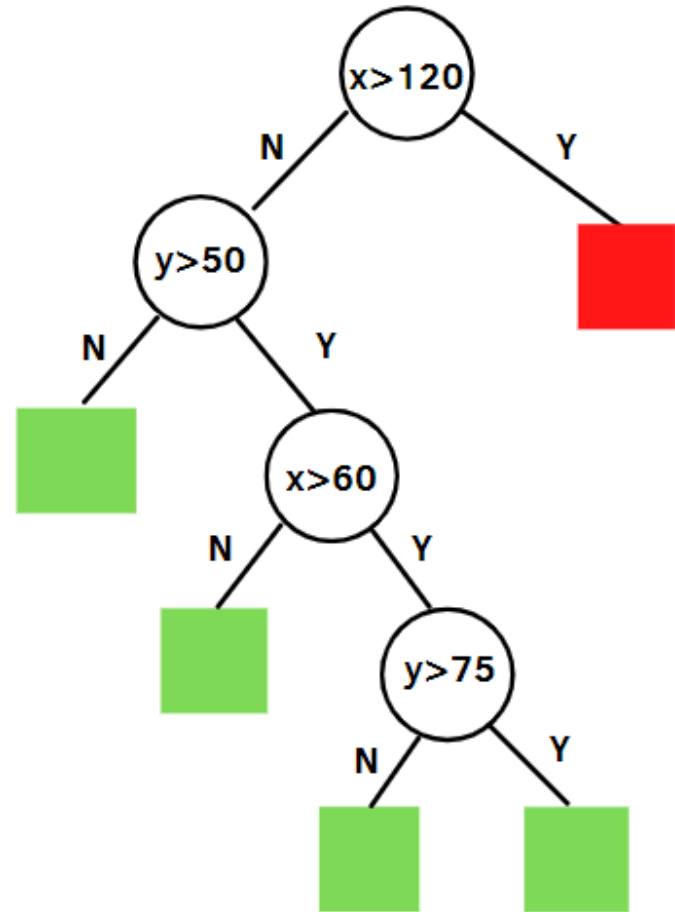
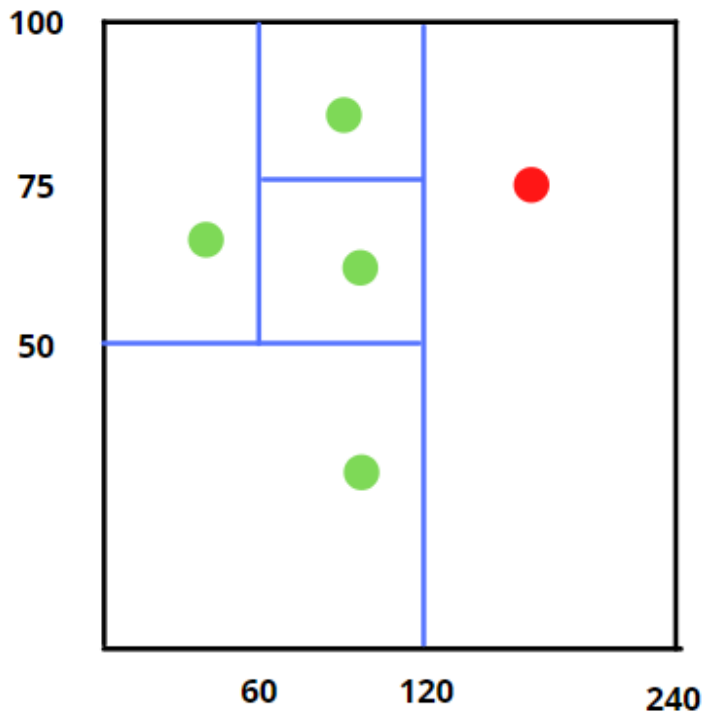
RD = Reachability Distance



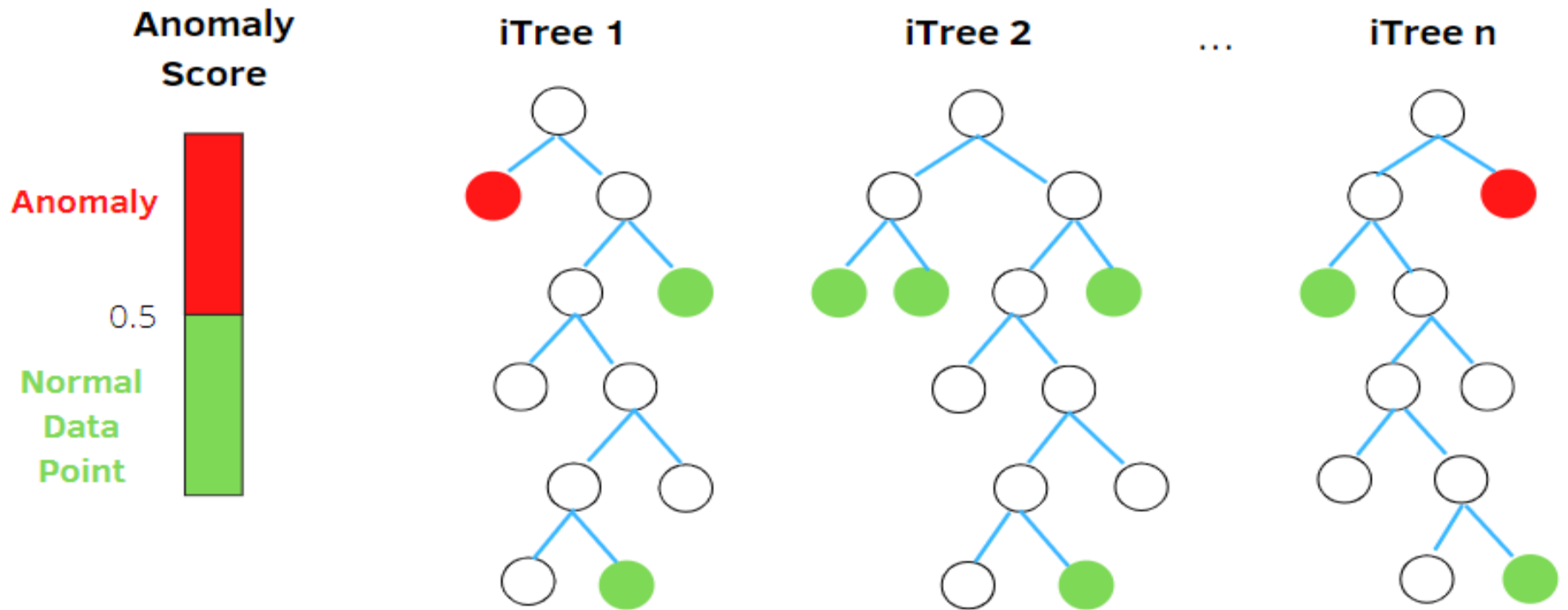
Calculate Local Reachability Distance for each data points



Изолирующий Лес Isolation Forrest

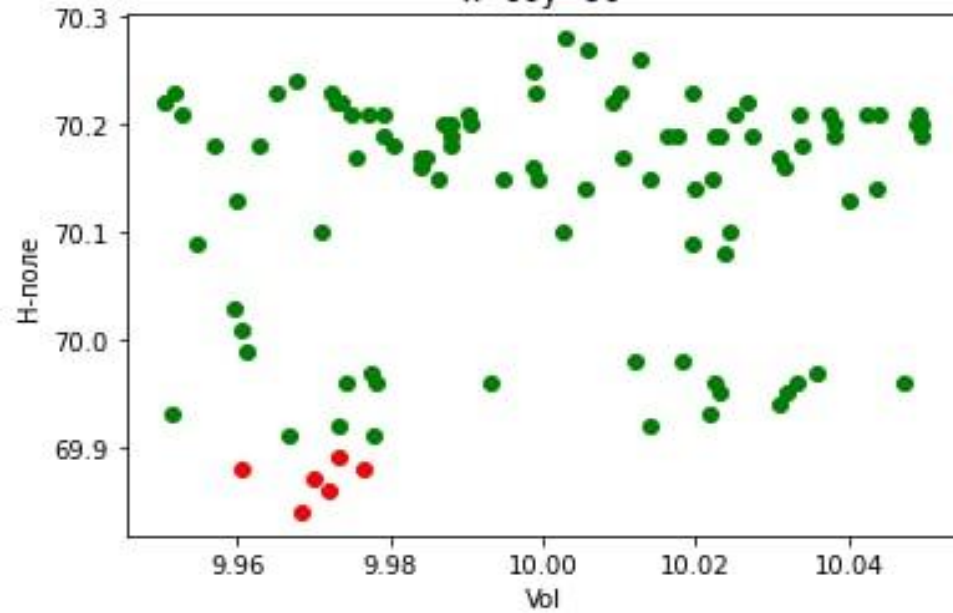


Isolation Forrest

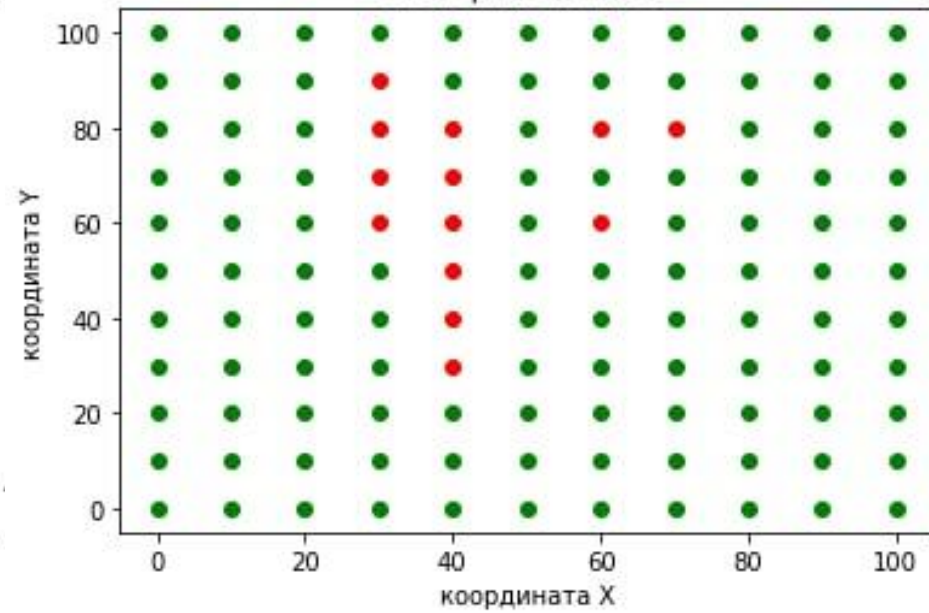


Эксперимент на реальных данных

$x=60, y=90$



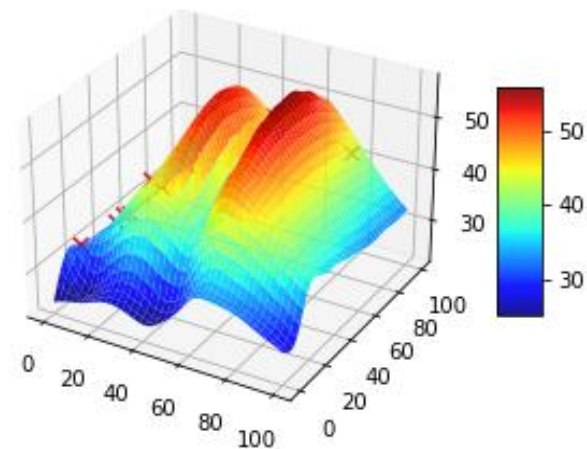
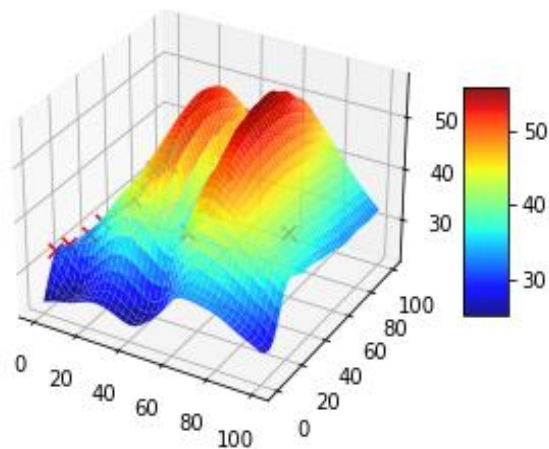
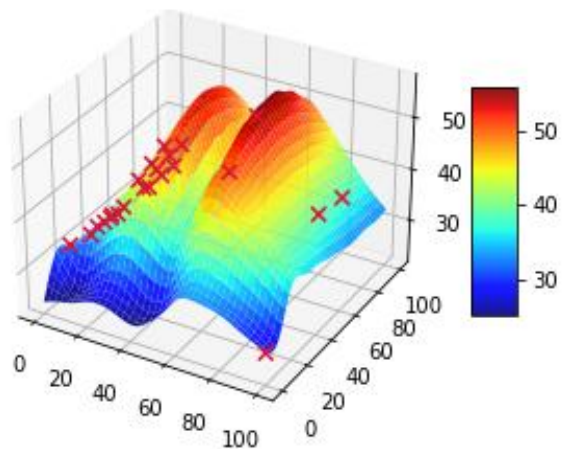
эксперимент № 83



Измерение 71 частота 100.0 МГц OCSVM

Измерение 71 частота 100.0 МГц __lof__

Измерение 71 частота 100.0 МГц iforest

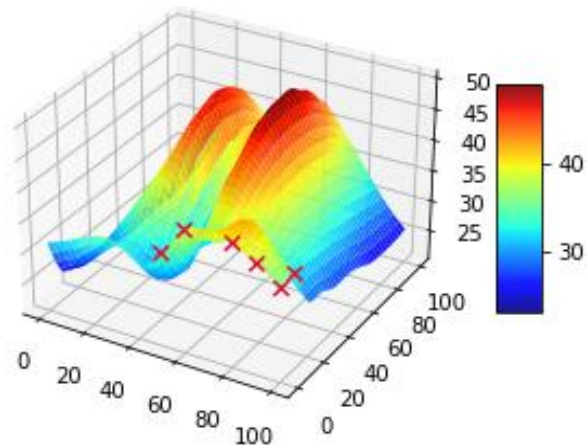
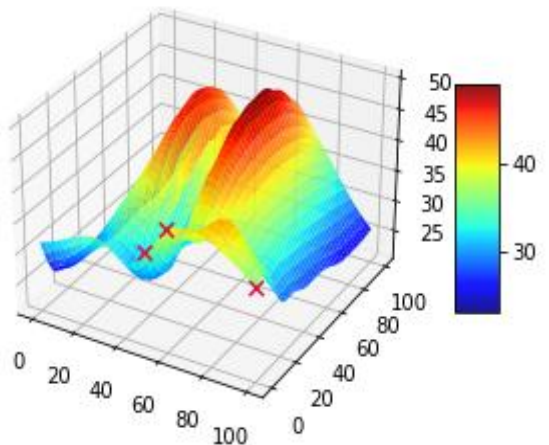
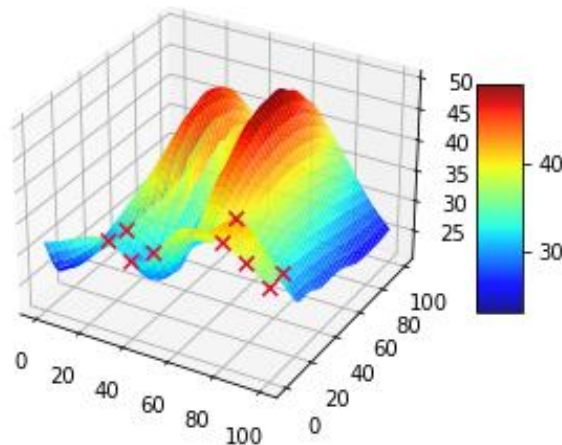


Проверка методов на экспериментальных данных

Измерение 84 частота 100.0 МГц OCSVM

Измерение 84 частота 100.0 МГц __lof__

Измерение 84 частота 100.0 МГц iforest



Оценка результатов

В качестве метрики для оценки работы алгоритмов применялась AUC ROC (Area Under Curve Receiver Operating Characteristic).

Samples	Dimensions	LOF	OCSVM	IForest
100	2	0,743	0,924	0,831

Интервал AUC	Качество модели
0,9 – 1,0	отличное
0,8-0,9	Очень хорошее
0,7-0,8	хорошее
0,6-0,7	среднее
0,5-0,6	неудовлетворительное

Выводы

Исходя из результатов проведенных экспериментах на реальных данных, а так же по результатам сравнительного анализа алгоритмов выявления аномалий можно выявить что для конкретного задачи лучшие результаты показал алгоритм OneClass – SVM.

Данный алгоритм может быть взят за основу прикладного программного обеспечения для контроля качества данных научного эксперимента в данной предметной области.

Список используемой литературы:

- 1) Arthur Zimek, Erich Schubert. Outlier Detection // Encyclopedia of Database Systems. — Springer New York, 2017.
- 2) 2) Борисьяк М. А. «Методы машинного обучения для контроля качества данных в научных экспериментах»: Дис. ... канд. тех. наук:15.10.20/ М. А. Борисьяк. - М., 2020. -116 с.
- 3) 3) А. М. Головина, А. Г. Дьяконов, «Выявление аномалий в работе механизмов методами машинного обучения»/Труды XIX Международной конференции «Аналитика и управление данными в областях с интенсивным использованием данных» (DAMDID/RCDL'2017), Москва, Россия, 10–13 октября 2017
- 4) 4) А. А. Шолохова, «Разработка методов обнаружения аномалий маневрирования и потребления топлива для морских судов» // Вып. квалиф. раб. аспиранта:02.06.01/ А. А. Шолохова. —СПБ., 2018. - 127 с.
- 5) 5) Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey // ACM Computing Surveys. — 2009. — Т. 41, вып. 3
- 6) 6) P. Guo, H. Kim, N. Virani, J. Xu, M. Zhu and P. Liu, "RoboADS: Anomaly Detection Against Sensor and Actuator Misbehaviors in Mobile Robots," 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2018, pp. 574-585.
- 7) 7) Mokhtari, S.; Abbaspour, A.; Yen, K.K.; Sargolzaei, A. A Machine Learning Approach for Anomaly Detection in Industrial Control Systems Based on Measurement Data. Electronics 2021, 10, 407
- 8) 8) Breunig, M. M.; Kriegel, H.-P.; Ng, R. T.; Sander, J. R. (1999). "OPTICS-OF: Identifying Local Outliers" Principles of Data Mining and Knowledge Discovery. Lecture Notes in Computer Science. Vol. 1704. p. 262
- 9) 9) Ramaswamy, Sridhar; Rastogi, Rajeev; Shim, Kyuseok (2000). "Efficient algorithms for mining outliers from large data sets". Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00. Proceedings of the 2000 ACM SIGMOD international conference on Management of data – SIGMOD '00. pp. 427–438.
- 10) 10) Mennatallah Amer, Markus Goldstein, Slim Abdennadher (2013). "Enhancing one-class Support Vector Machines for unsupervised anomaly detection". ODD'13: Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description, pp. 8-15.
- 11) 11) Ester, Martin; Kriegel, Hans-Peter; Sander, Jörg; Xu, Xiaowei (1996). Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M. (eds.). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231.
- 12) 12) Hinton GE, Krizhevsky A, Wang SD. Transforming auto-encoders. In International Conference on Artificial Neural Networks 2011 Jun 14 (pp. 44-51). Springer, Berlin, Heidelberg.
- 13) 13) Cussens J (2011). "Bayesian network learning with cutting planes". Proceedings of the 27th Conference Annual Conference on Uncertainty in Artificial Intelligence: 153–160.
- 14) 14) Stuart J. Russell, Peter Norvig (2010) Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall
- 15) 15) Zhu, Xiaojin (2008). "Semi-supervised learning literature survey" . University of Wisconsin-Madison.

**Спасибо
за внимание!**

23.01.23