

# Поиск аномалий

iForest и PCA

# Обо мне

- Старший специалист по машинному обучению
- deep learning engineer
- NLP, CV, anomaly detection
- Open source contributor
- Выпускник и амбассадор Яндекс Практикума
- Выпускник DLS ФПМИ МФТИ



# Аномалии

# Применение



# Применение



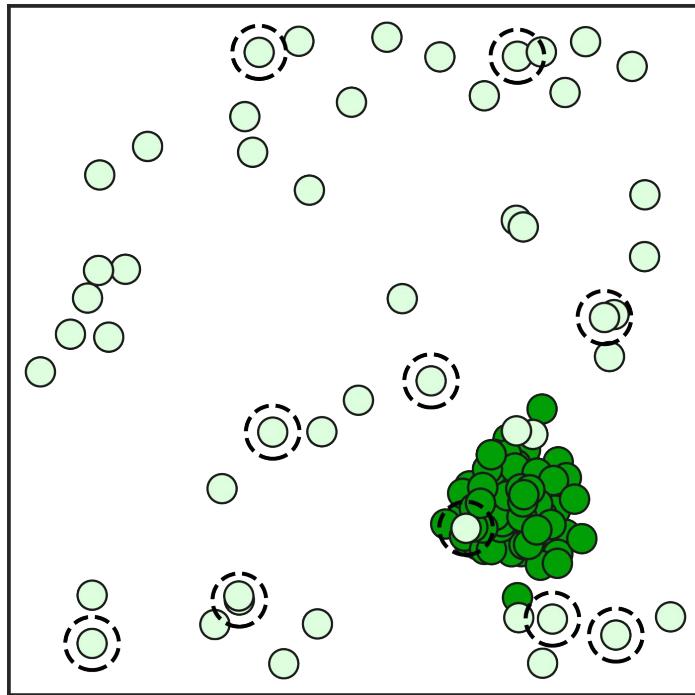
# Применение



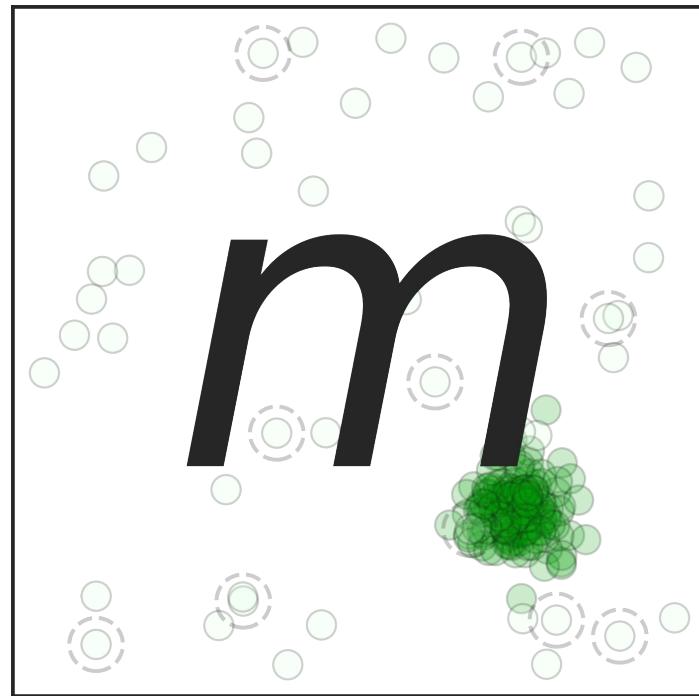
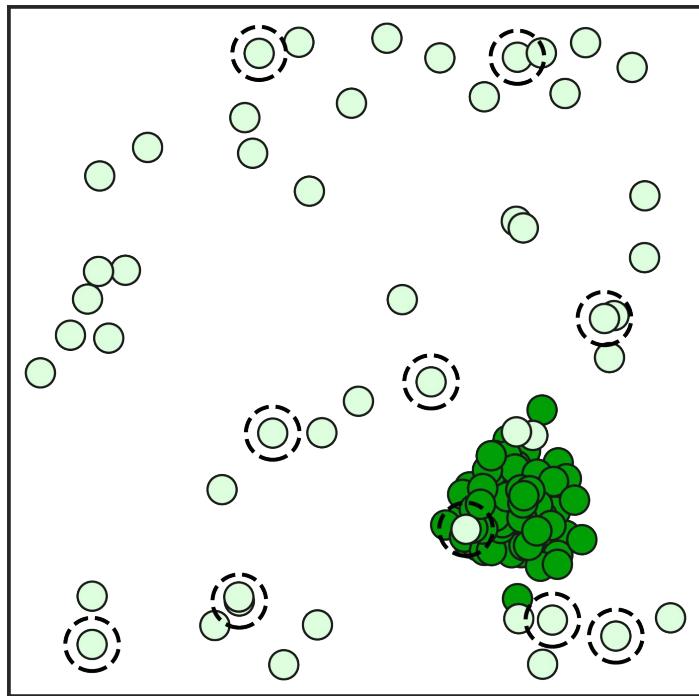
# Применение



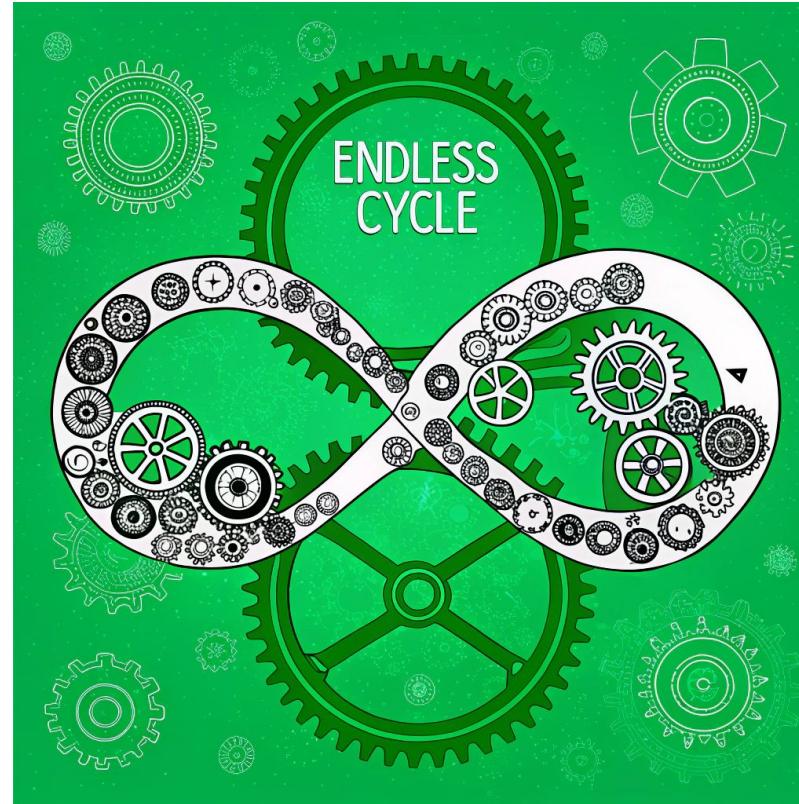
# Этапы



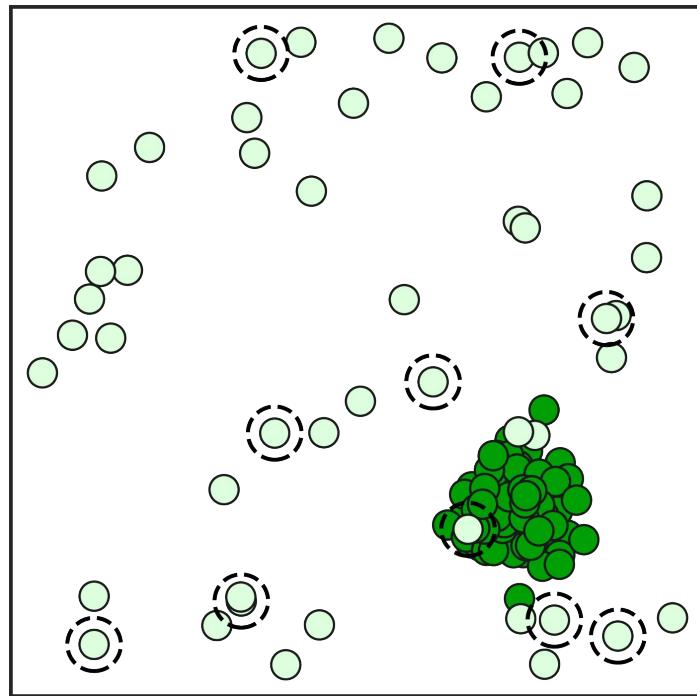
# Этапы



# Применение



# Сбор данных



<b>попугаи</b>	<b>удавы</b>
3.248357	-2.874754
2.930868	-2.826776
3.323844	-3.340012
3.761515	-2.883873
2.882923	-2.853464
2.882932	-3.357176
3.789606	-2.067113
...	...

<b>попугаи</b>	<b>удавы</b>
3.248357	-2.874754
2.930868	-2.826776
3.323844	-3.340012
3.761515	-2.883873
2.882923	-2.853464
2.882932	-3.357176
3.789606	-2.067113
...	...

<b>индекс аномальности</b>
-0.119150
-0.131275
-0.108670
-0.065472
-0.128972
-0.120056
-0.012170
...

# Методы

```
$ pip install pyod  
from pyod.models.iforest import IForest  
clf = IForest()  
clf.fit(data)  
scores = clf.decision_scores_
```

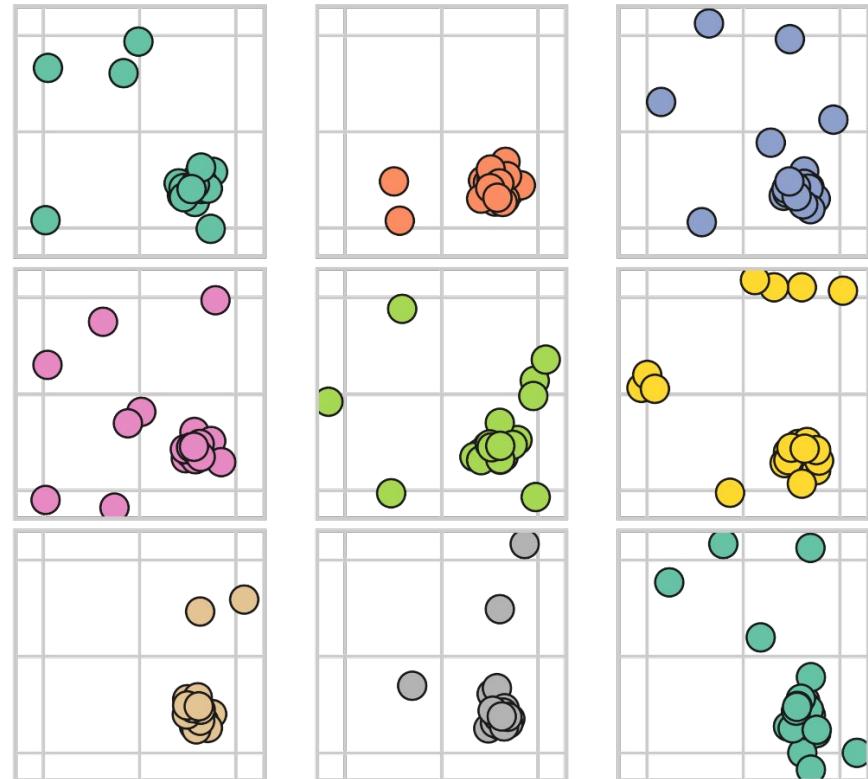
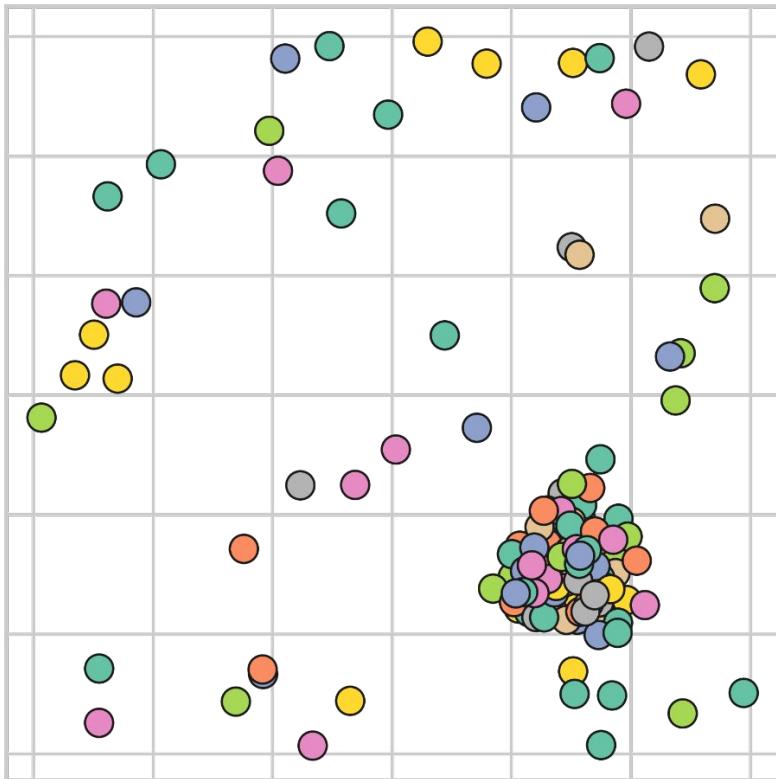
# Методы

```
$ pip install pyod
```

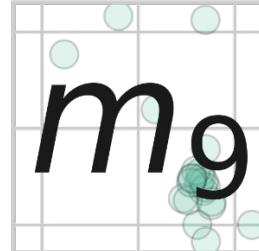
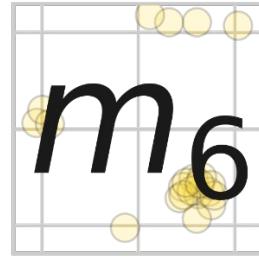
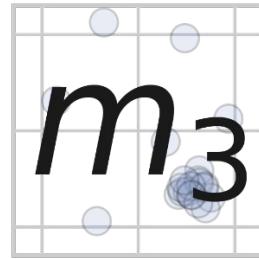
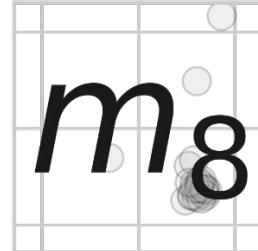
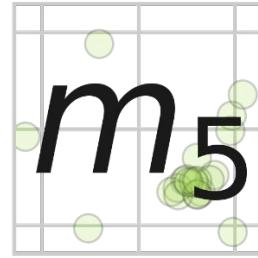
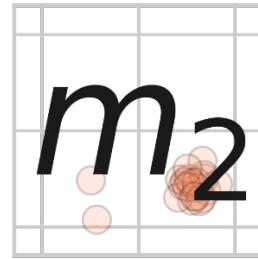
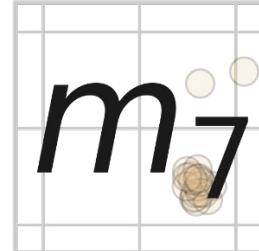
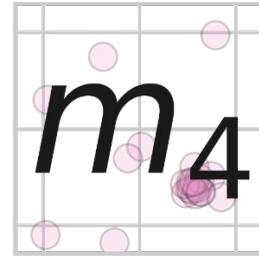
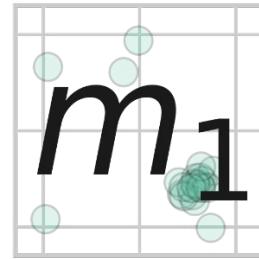
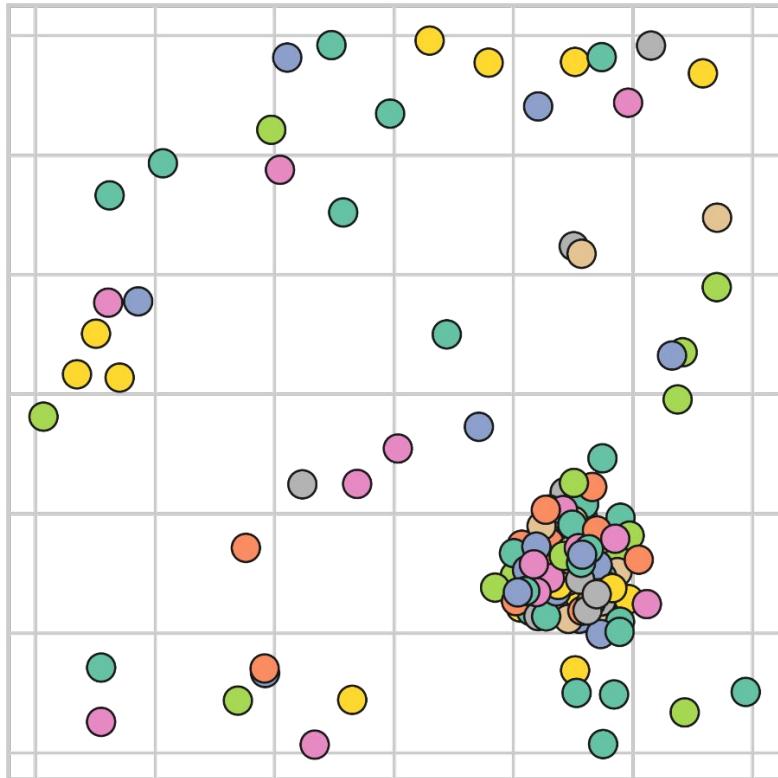
- [1] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, 413–422. IEEE, 2008.
- [2] Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. “Isolation-based anomaly detection.” *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1 (2012): 3.
- [3] Mei-Ling Shyu, Shu-Ching Chen, Kanoksri Sarinnapakorn, and LiWu Chang. A novel anomaly detection scheme based on principal component classifier. Technical Report, MIAMI UNIV CORAL GABLES FL DEPT OF ELECTRICAL AND COMPUTER ENGINEERING, 2003.

# Ансамбли

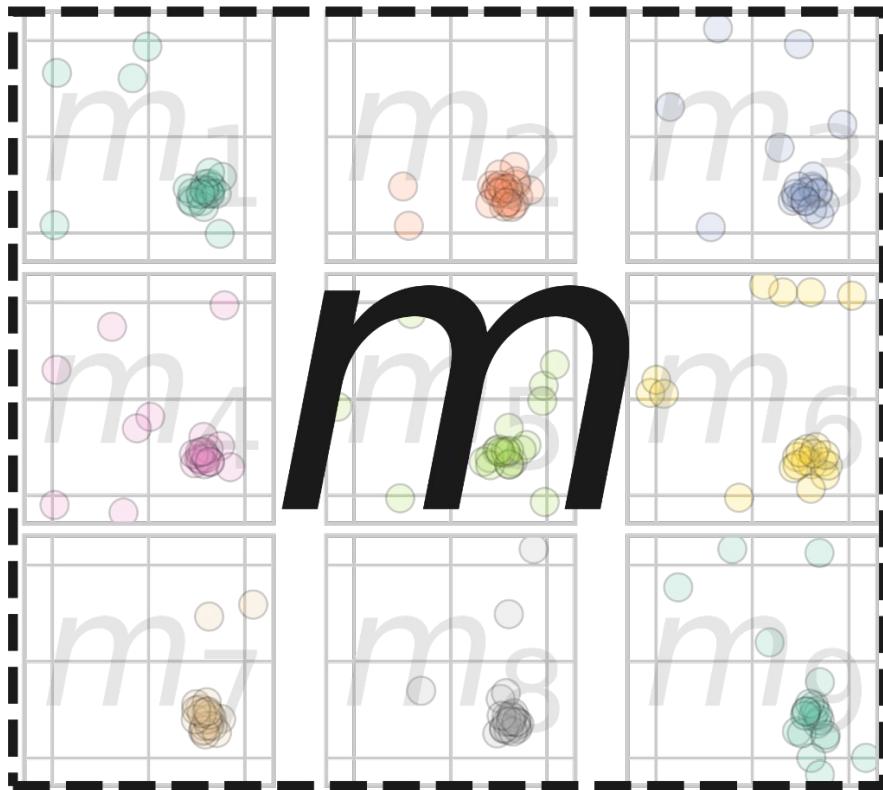
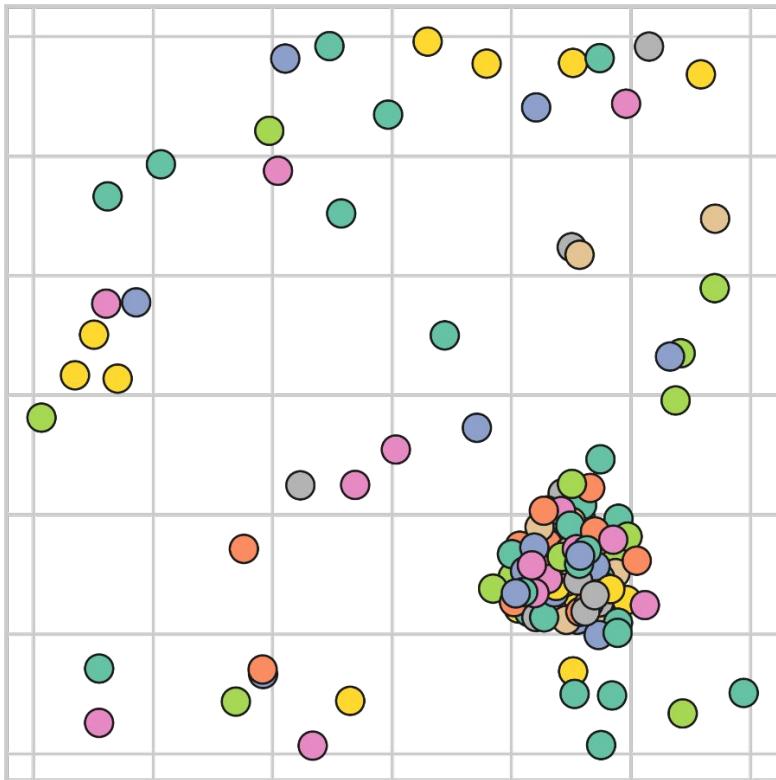
# Делим



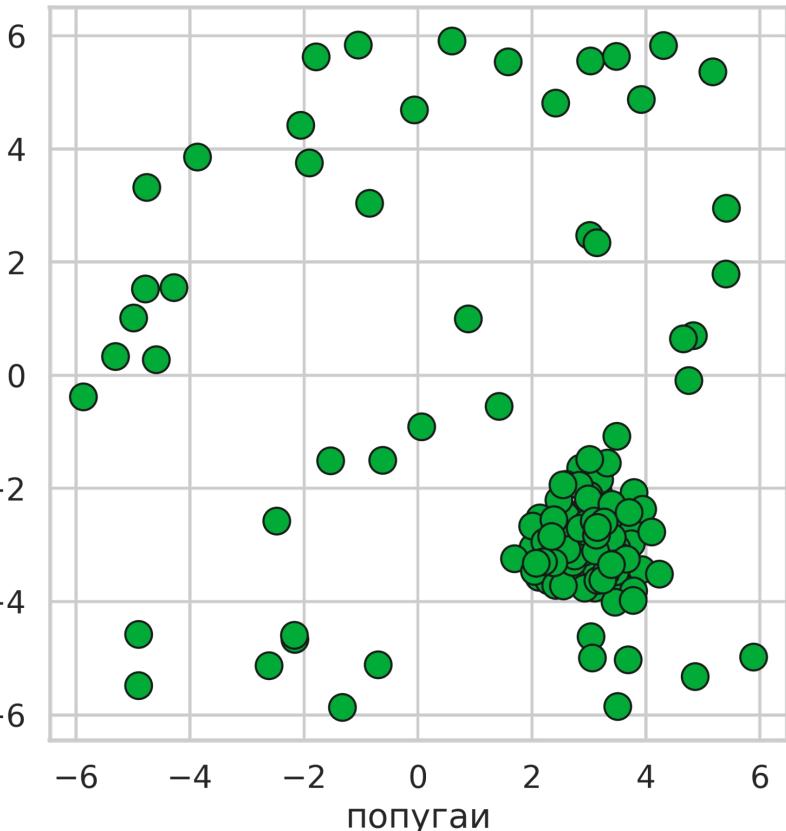
# Обучаем



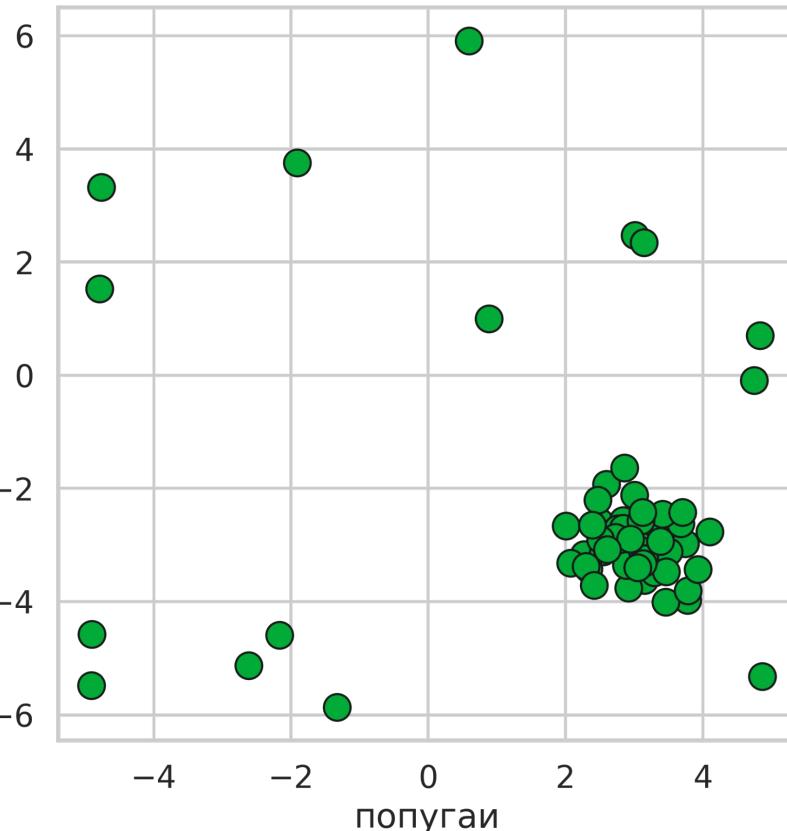
# Усредняем



удавы



удавы



In [1]: import random

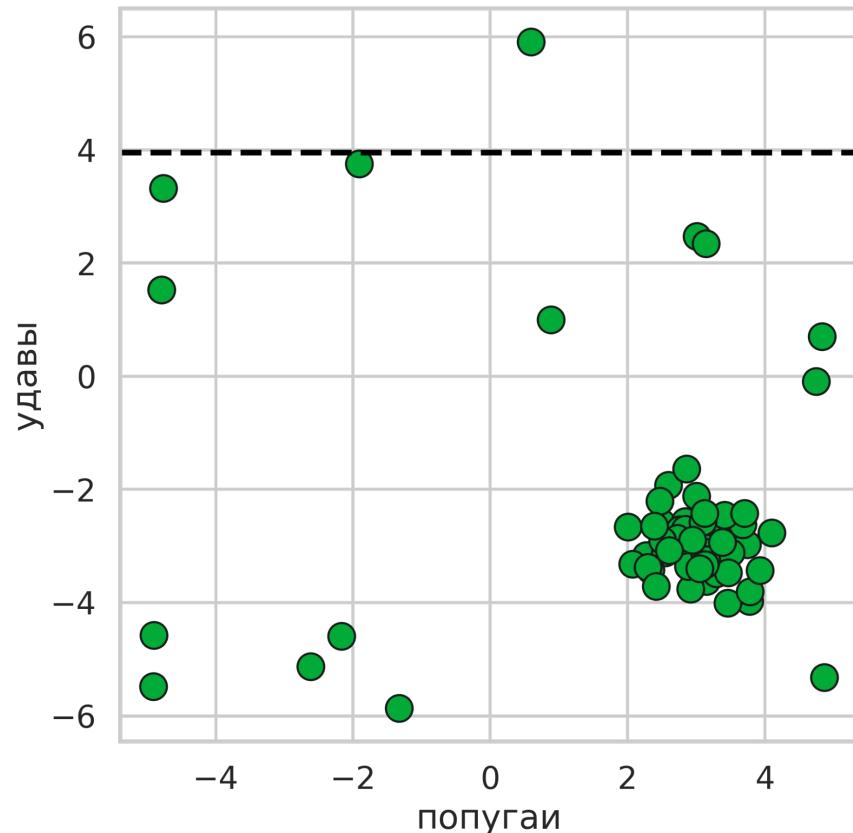
In [2]: axes = ['попугай', 'удавы']

In [3]: random.choice(axes)

Out[3]: 'удавы'

In [4]: random.uniform(-6.0, 6.0)

Out[4]: 3.954197818641566



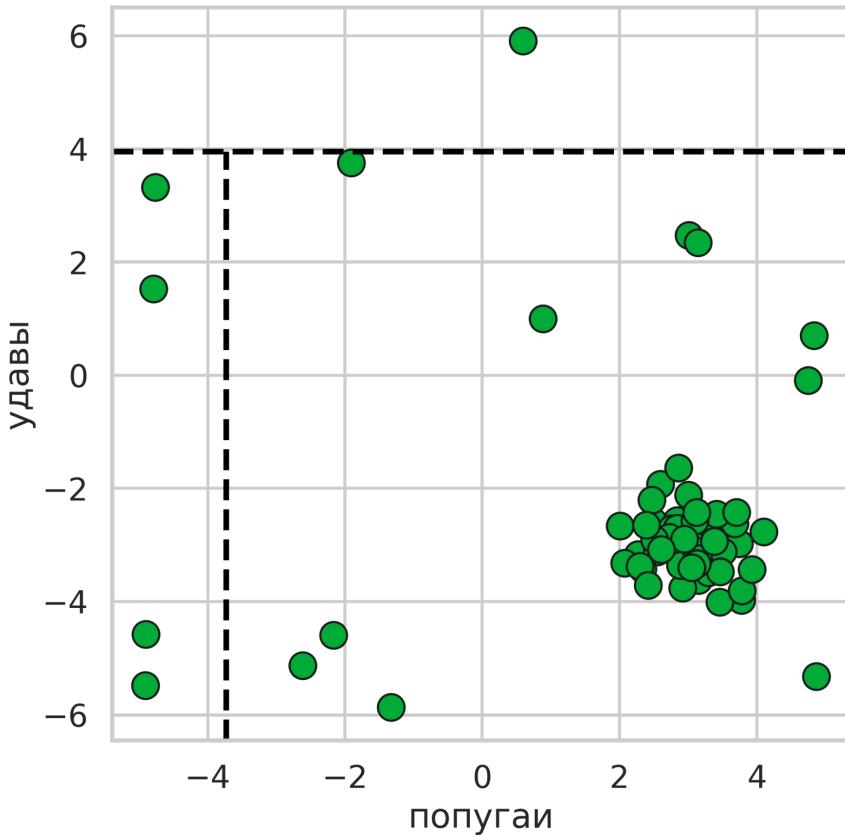
...

In [5]: `random.choice(axes)`

Out[5]: 'попугай'

In [6]: `random.uniform(-6.0, 3.95)`

Out[6]: -3.7345546743319455



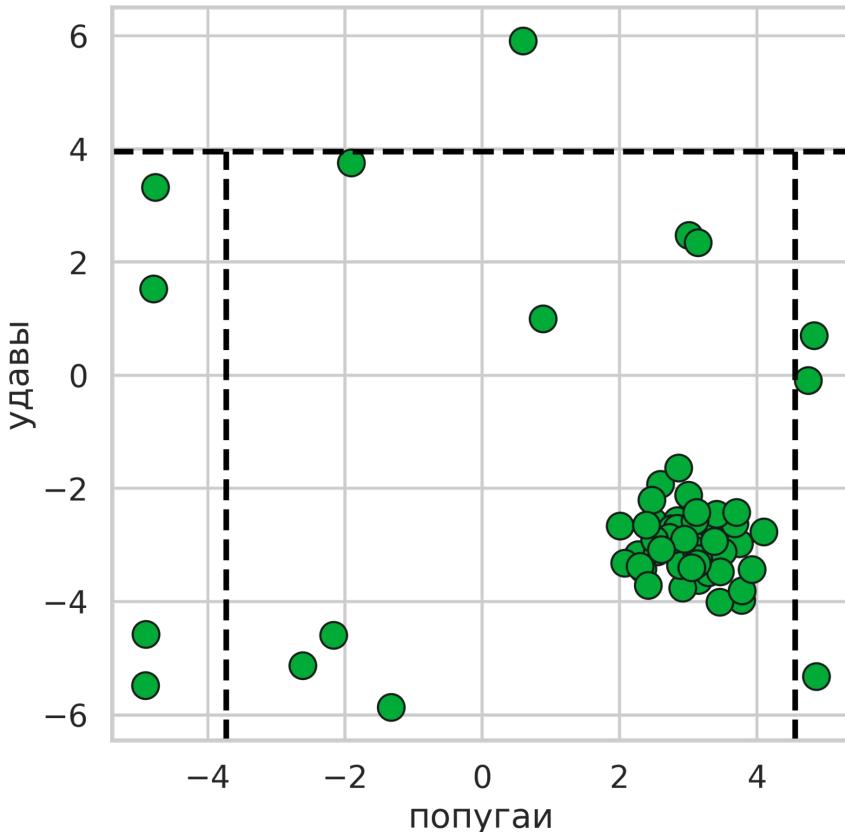
...

In [7]: `random.choice(axes)`

Out[7]: 'попугай'

In [8]: `random.uniform(-3.73, 6.0)`

Out[8]: 4.55352143693694



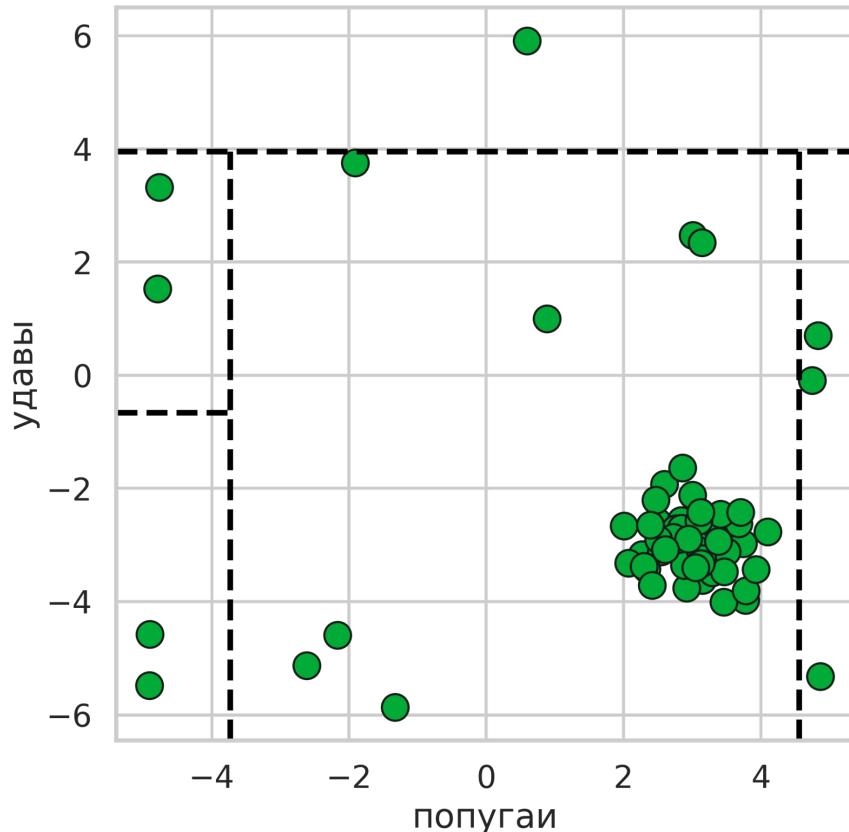
...

In [9]: `random.choice(axes)`

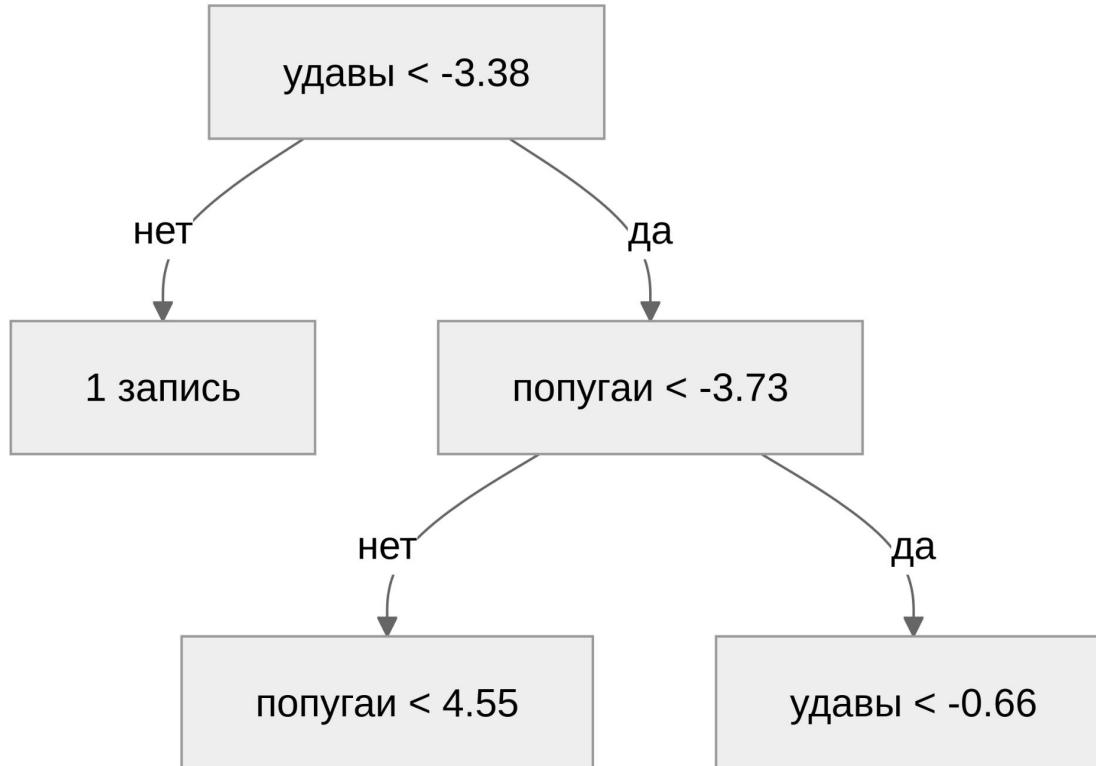
Out[9]: 'удавы'

In [10]: `random.uniform(-6.00, -3.73)`

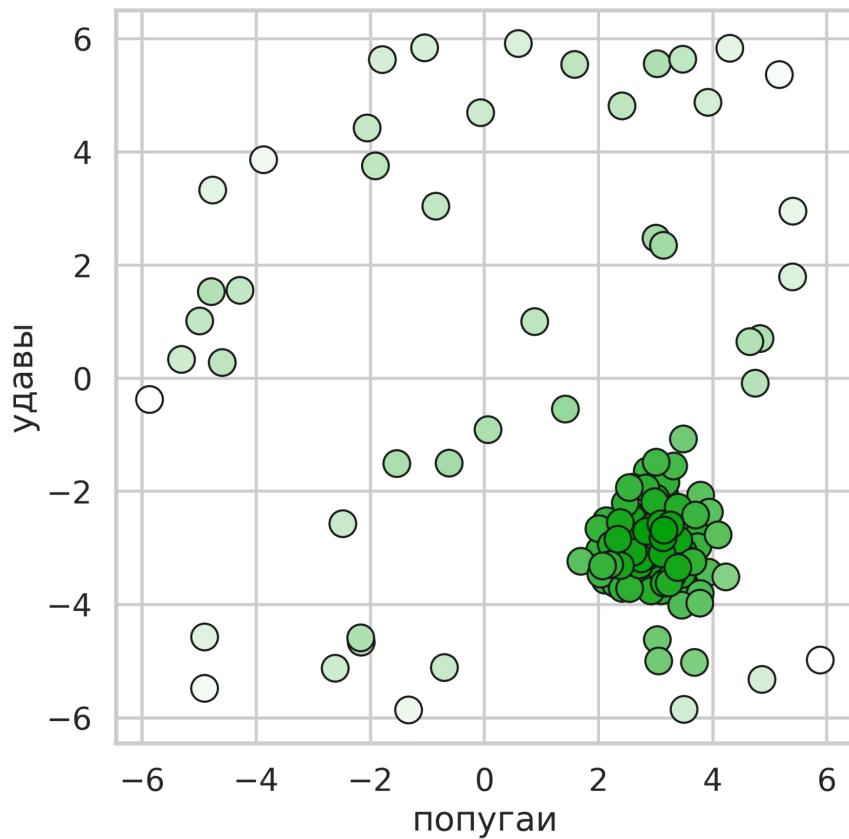
Out[10]: -0.6572094533790986



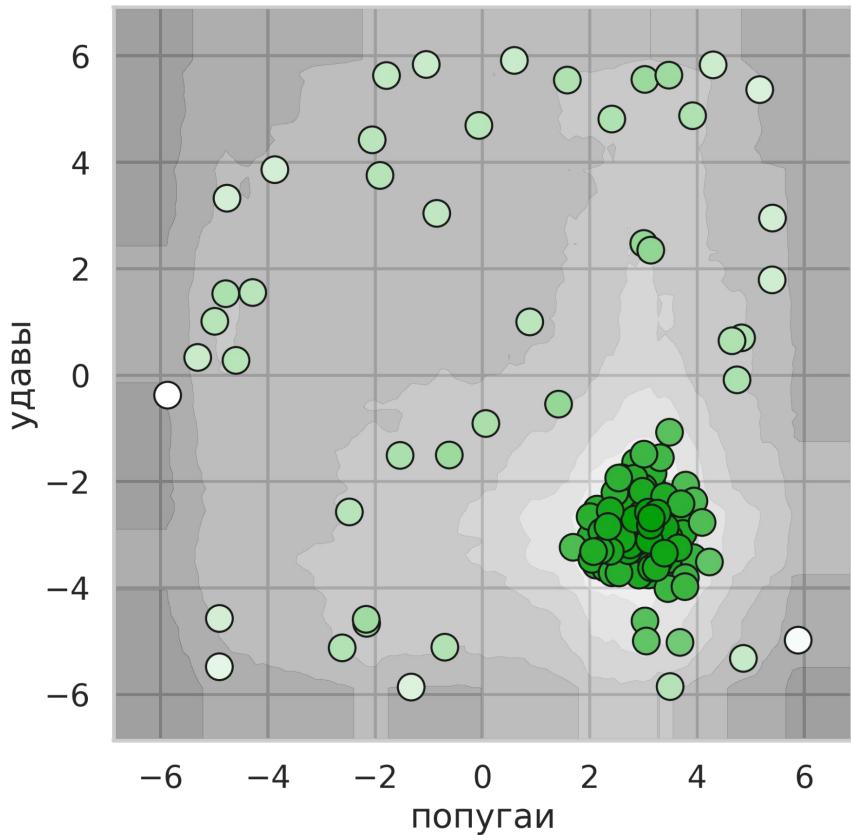
# iTree

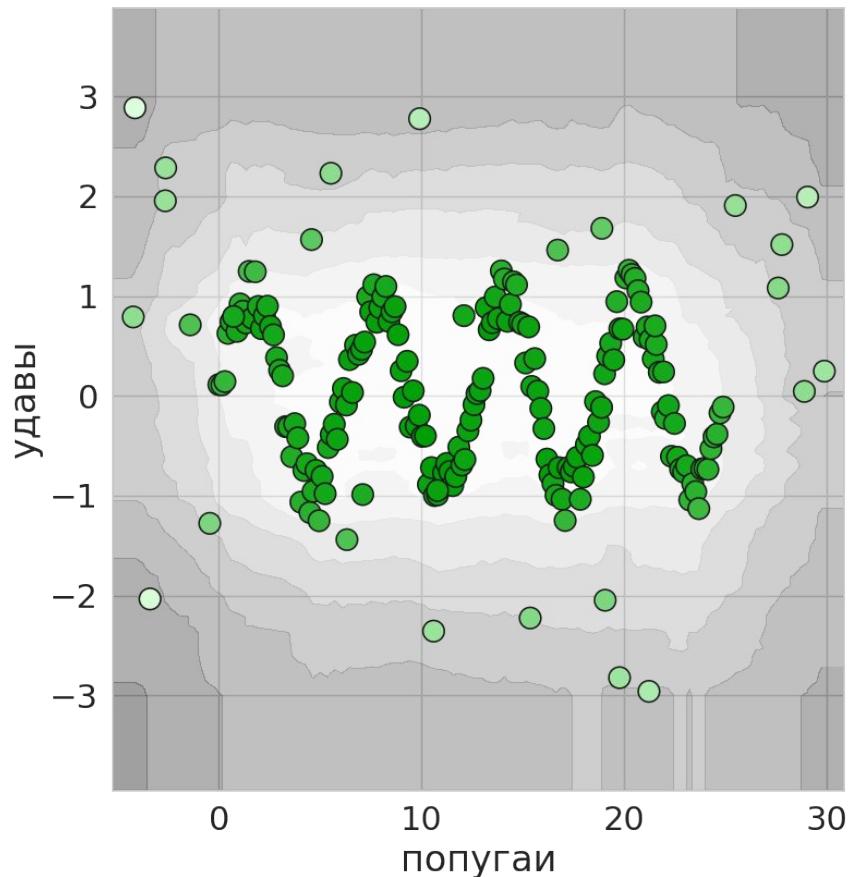


# Isolation Forest [1][2]

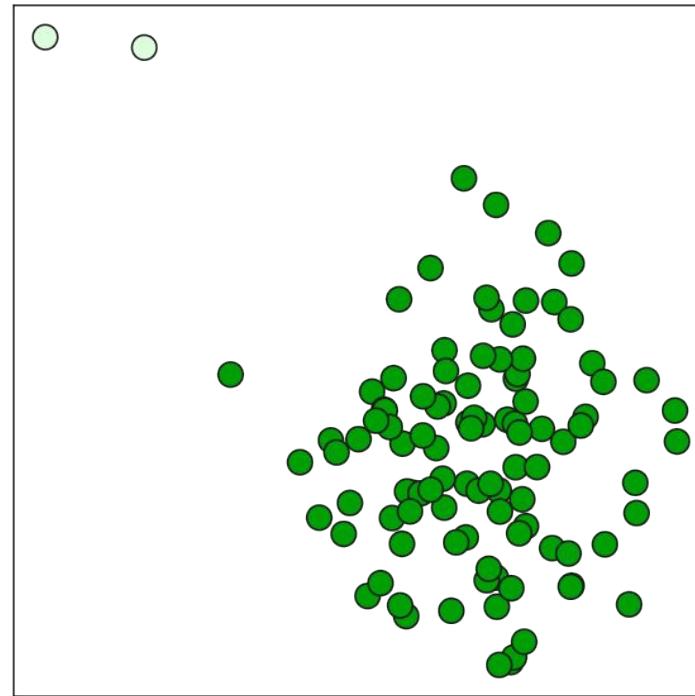


# Isolation Forest [1][2]

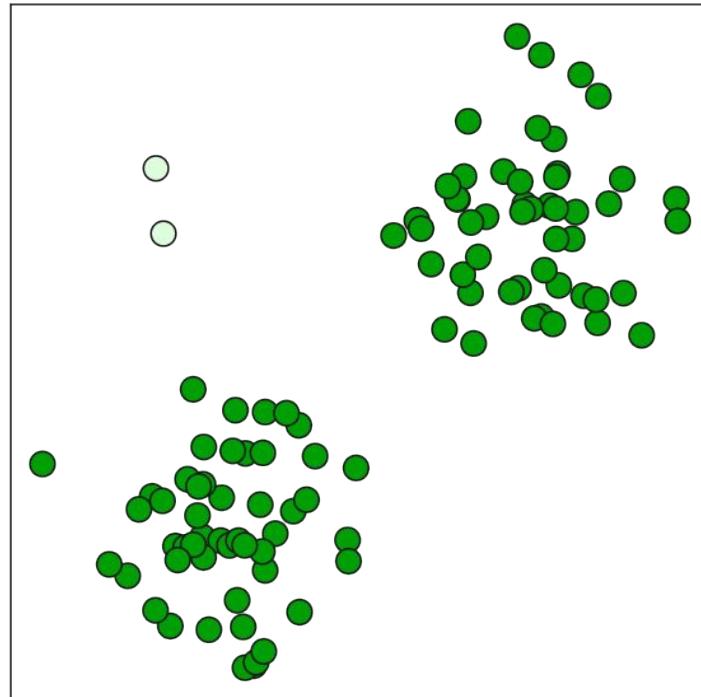


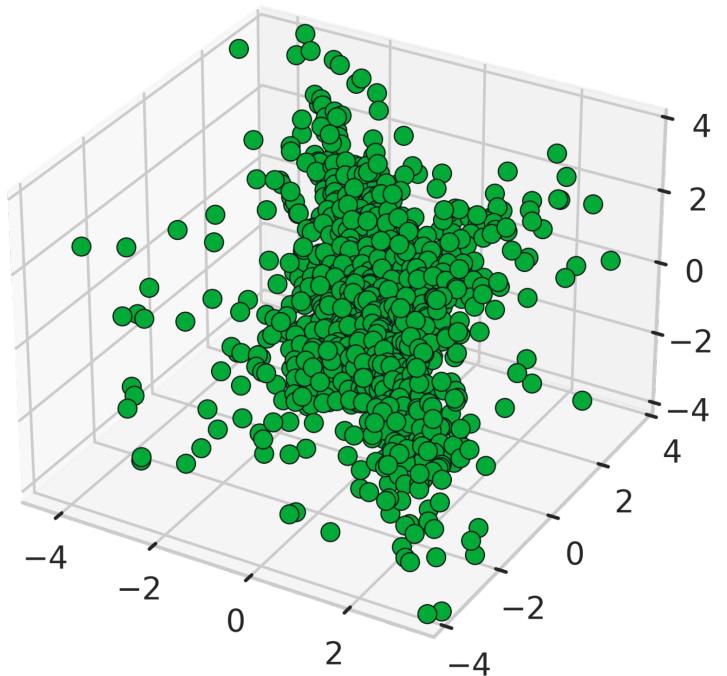


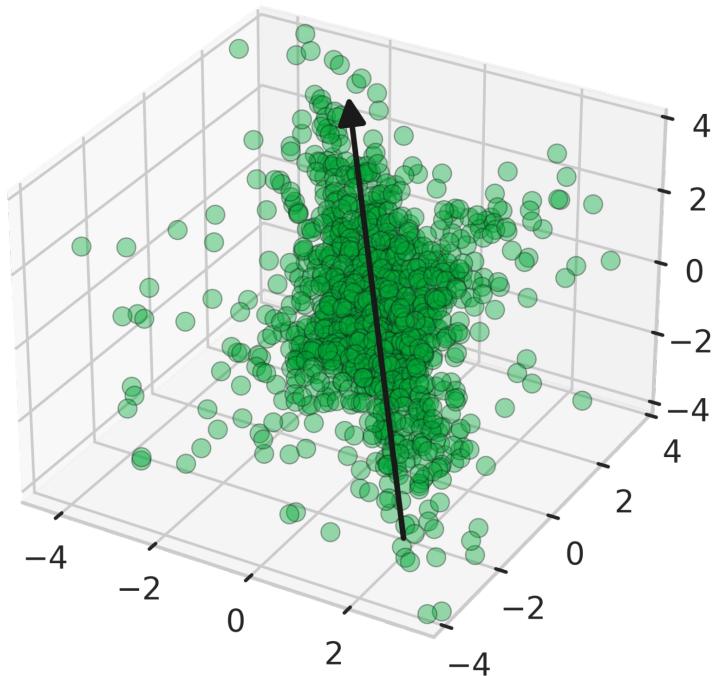
# Глобальные

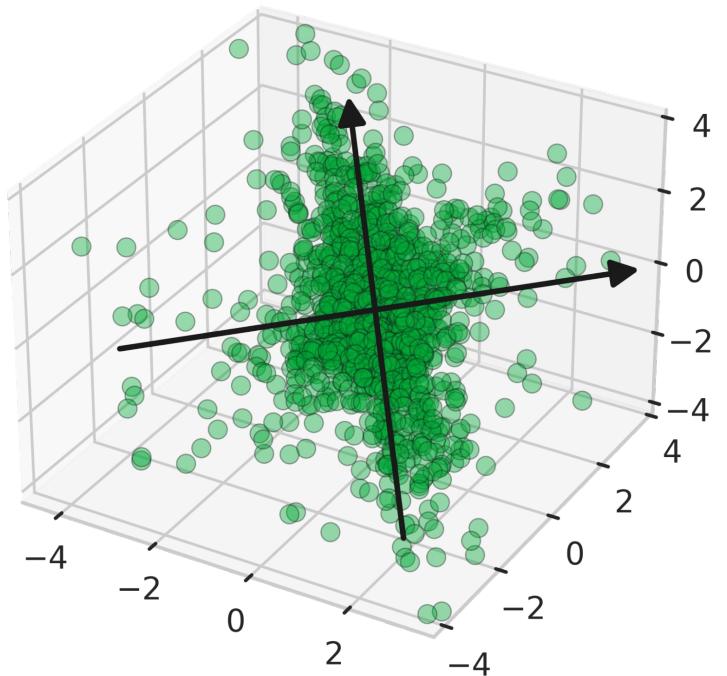


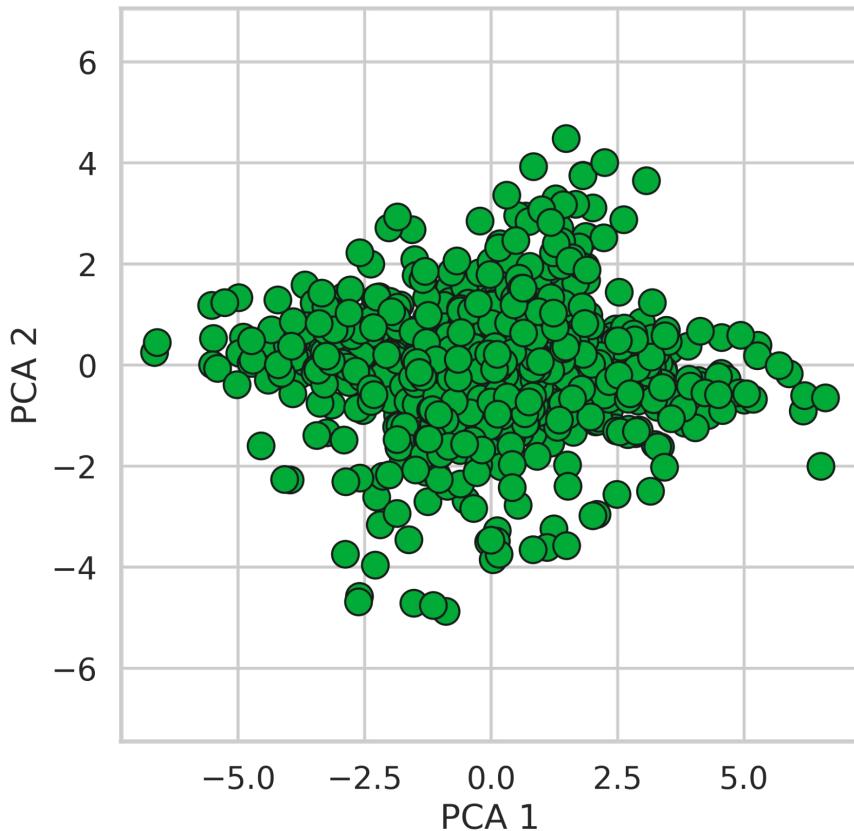
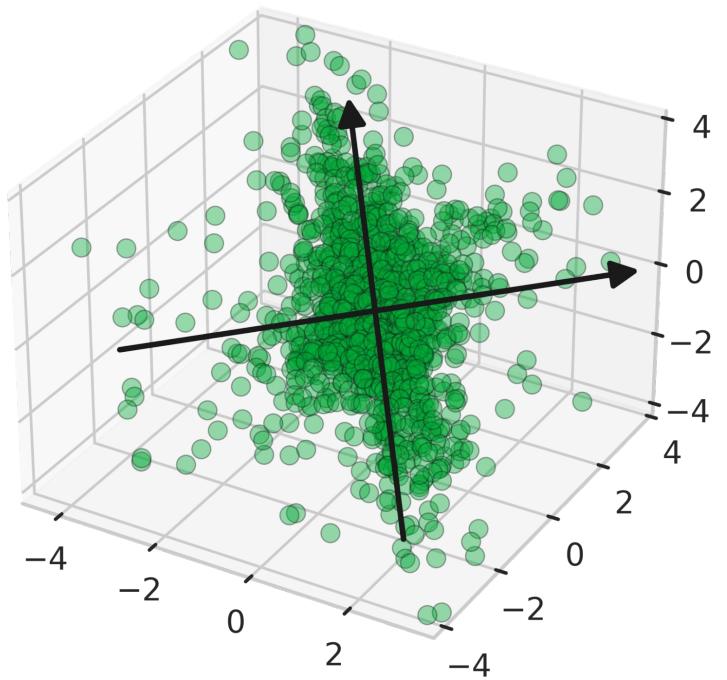
# Контекстные

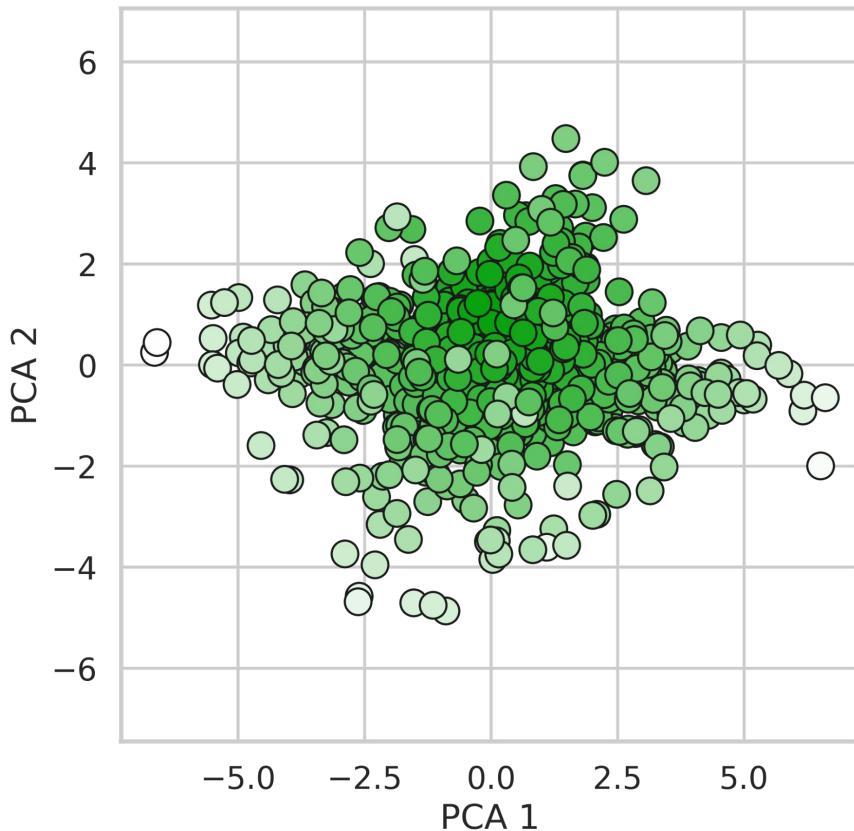
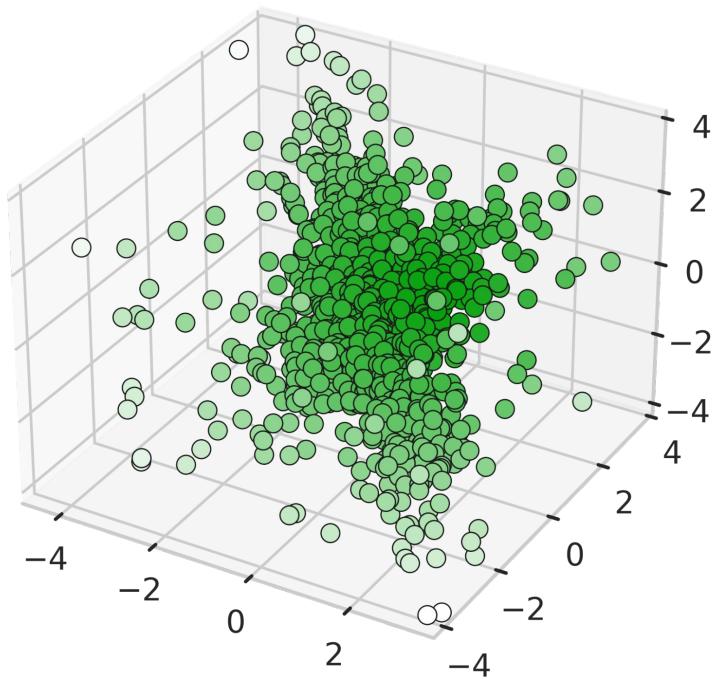




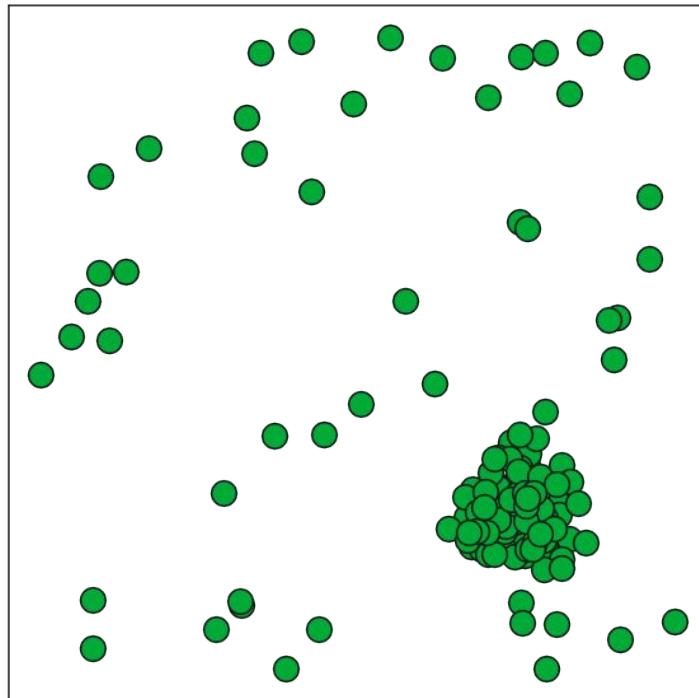




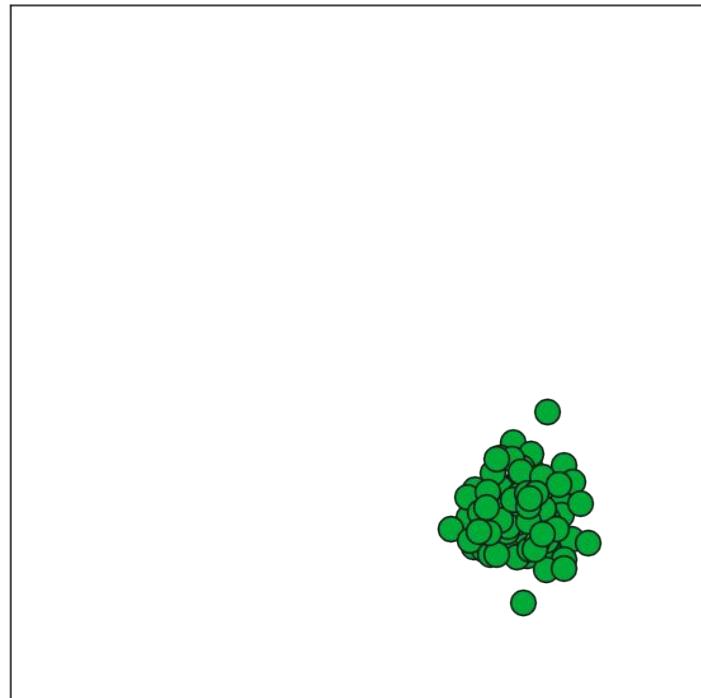
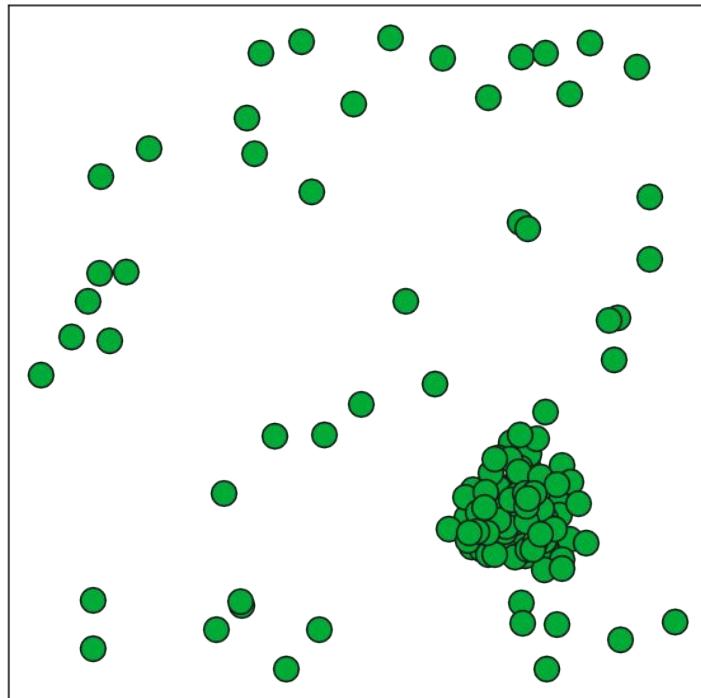




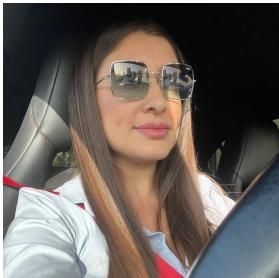
# Виды задач



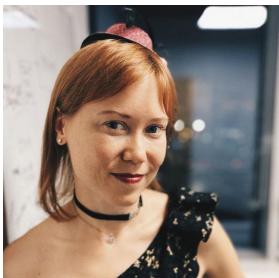
# Виды задач



# Резюме



# Котики



Лужники — 24 мая, 15:00

## Поиск аномалий с использованием Python: от теории к практике

Михаил Васильев

Makves



Moscow Python

