

Поиск аномалий

Получение списка выбросов из колонки column

```
In boxplot = plt.boxplot(df['column'].values)
outliers = list(boxplot["fliers"][0].get_data()[1]) print(model.inertia_)
```

Обучение изоляционного леса для поиска аномалий

```
from sklearn.ensemble import IsolationForest
isolation_forest = IsolationForest(n_estimators=100)
isolation_forest.fit(data)

In # получение оценки аномальности.
# Значения от -0.5 до 0.5. Чем ниже, тем вероятнее, что объект – аномалия
anomaly_scores = isolation_forest.decision_function(data)

# получение предсказания аномальности. -1 – выброс, 1 – нормальный объект
estimator = isolation_forest.predict(data)

# обучение и получение предсказаний
estimator = isolation_forest.fit_predict(data)
```

Поиск аномалий методом ближайших соседей

```
In from pyod.models.knn import KNN
model = KNN()
model.fit(data)

# предсказания аномальности: 1 – аномалия, 0 – нормальный объект
predictions = model.predict(data)
```

Словарь

Аномалии, или выбросы (anomalies; outliers)

объекты с «ненормальным» поведением, то есть отклоняющимся от общего тренда. Они указывают на проблему в данных или на нестандартные события в данных.