

Датасэты для задач машинного обучения

1. Kaggle

Около 30 000 датасэтов по различной тематике и для решения различных задач

<https://www.kaggle.com/datasets>

2. UCI Machine Learning Repository

Около 500 датасэтов

<https://archive.ics.uci.edu/ml/index.php>

3. Scikit-learn

<https://scikit-learn.org/stable/datasets/index.html>

4. Drivendata.org

Соревнования по машинному обучению для решения социальных проблем

<https://www.drivendata.org/competitions/>

5. Visualdata.io

Более 400 датасэтов для решения задач, связанных с обработкой изображений, компьютерным зрением или глубоким обучением

<https://www.visualdata.io/>

6. Microsoft, Amazon, Google

<https://msropendata.com/>

<https://registry.opendata.aws/>

<https://datasetsearch.research.google.com/>

7. Awesome Public Datasets

<https://github.com/awesomedata/awesome-public-datasets>

8. Финансовые и экономические данные

<https://finance.yahoo.com/>

<https://www.quandl.com/>

<https://markets.ft.com/data/archive>

9. Данные государственных органов и международных организаций

Организация Объединенных Наций

<https://data.un.org/>

Международный валютный фонд

<https://data.imf.org/datasets>

Всемирный банк

<https://data.worldbank.org/>

Центральный банк России

<https://www.cbr.ru/>

Открытые данные Правительства США

<https://www.data.gov/>