ELITE DATA SCIENCE

DATASETS FOR DATA SCIENCE AND MACHINE LEARNING

Here are some of our favorite datasets for DIY data science and machine learning projects. They are broken into categories. To see the latest version, plus detailed annotations, visit on the online dataset list at EliteDataScience.com.

1. EXPLORATORY ANALYSIS

OUR PICKS:
Game of Thrones
World University Rankings
IMDB 5000 Movie Dataset

AGGREGATORS:
Kaggle Datasets
r/datasets

2. GENERAL MACHINE LEARNING

OUR PICKS:
Wine Quality (Regression)
Credit Card Default (Classification)
US Census Data (Clustering)

AGGREGATORS:
UCI Machine Learning Repository

3. DEEP LEARNING

OUR PICKS:
MNIST
CIFAR
ImageNet
YouTube 8M

AGGREGATORS:
Deeplearning.net
DeepLearning4J.org

4. NATURAL LANGUAGE PROCESSING

OUR PICKS:
Enron Dataset
Amazon Reviews
Newsgroup Classification

AGGREGATORS:
nlp-datasets (Github)
Quora Answer

5. CLOUD MACHINE LEARNING

OUR PICKS:
AWS Public Datasets
Google Cloud Public Datasets
Microsoft Azure Public Datasets

ELITEDATASCIENCE.COM

DATASETS FOR DATA SCIENCE AND MACHINE LEARNING

6. TIME SERIES ANALYSIS

OUR PICKS:	AGGREGATORS:
EOD Stock Prices	Quandl
Zillow Real Estate Research	The World Bank
Global Education Statistics	

7. RECOMMENDER SYSTEMS

OUR PICKS:	AGGREGATORS:
MovieLens	entaroadun (Github)
Jester	
Million Song Dataset	

8. SPECIFIC INDUSTRIES

OUR PICKS:
Awesome Public Datasets
Data.gov

9. STREAMING

OUR PICKS:	AGGREGATORS:
Twitter API	Satori
StockTwits API	
Weather Underground	

10. WEB SCRAPING

OUR PICKS:
ToScrape.com

11. CURRENT EVENTS

AGGREGATORS:
FiveThirtyEight
BuzzFeedNews

To see the latest version, plus detailed annotations, visit on the online dataset list at EliteDataScience.com.