

Data Collection for ML

July 11, 2024

TOPICS

- Importance of Data in Machine Learning?
- Using machine learning model to differentiate one item from the other.
- Where to collect the Data? Kaggle UCI Machine Learning Repository Google Dataset Search
- Demonstration of Data Collection
- Importance of Data in Machine Learning

APPLICATION PROGRAMMING INTERFACE (API)

Importing Datasets through Kaggle API

```
[4]: import warnings

# Ignore specific warnings related to the installation
warnings.filterwarnings("ignore", category=UserWarning)
warnings.filterwarnings("ignore", category=DeprecationWarning)

# Install the kaggle library without the verbose output
!pip install --quiet kaggle

import warnings
```

```
[7]: ! kaggle datasets list -s "phone"
```

Traceback (most recent call last):

```
File "<frozen runpy>", line 198, in _run_module_as_main
File "<frozen runpy>", line 88, in _run_code
File "C:\Users\DONATUS\anaconda24\Scripts\kaggle.exe\__main__.py", line 4, in
<module>
File "C:\Users\DONATUS\anaconda24\Lib\site-packages\kaggle\__init__.py", line
7, in <module>
    api.authenticate()
File "C:\Users\DONATUS\anaconda24\Lib\site-
packages\kaggle\api\kaggle_api_extended.py", line 398, in authenticate
    raise IOError('Could not find {}'. Make sure it's located in'
OSError: Could not find kaggle.json. Make sure it's located in
C:\Users\DONATUS\.kaggle. Or use the environment method.
```

```
[8]: !kaggle datasets list -s "phone"
```

ref	size	lastUpdated	downloadCount	voteCount	usabilityRating	title
berkayeserr/phone-prices	27KB	2023-09-06 15:09:29	3265	44	0.88235295	Phone Prices
jacksondivakarr/phone-classification-dataset	205KB	2023-12-12 15:21:07	5014	73	1.0	Phone Classification Dataset
rkiattisak/mobile-phone-price	5KB	2023-03-09 04:15:24	6417	68	1.0	Mobile Phone Price
atefehmirnaseri/cell-phone-price	71KB	2024-04-19 11:16:35	3185	43	1.0	Cell Phone Price
grikomsn/amazon-cell-phones-reviews	9MB	2019-12-26 22:21:16	19055	224	1.0	Amazon Cell Phones Reviews
muhammedtausif/best-selling-mobile-phones	5KB	2022-10-08 11:57:44	3572	116	0.9117647	Mobile Phones
PromptCloudHQ/amazon-reviews-unlocked-mobile-phones	33MB	2017-01-11 10:22:30	15307	177	1.0	Amazon Reviews: Unlocked Mobile Phones
prasertk/mobile-phone-rating	3KB	2022-02-13 08:02:17	2043	26	1.0	Mobile phone rating
arwinneil/gsmarena-phone-dataset	635KB	2017-06-29 07:09:31	6514	85	0.85294116	GSMarena Phone Dataset
marcodena/mobile-phone-activity	242MB	2019-11-14 06:11:32	14163	272	0.7058824	Mobile phone activity in a city
dewangmoghe/mobile-phone-price-prediction	40KB	2024-05-20 14:32:40	2761	33	0.9411765	Mobile Phone Price Prediction
redpen12/mobile-phone-market-in-ghana	49KB	2023-10-18 10:33:10	781	67	1.0	Mobile Phone Market in Ghana
artempozdniakov/ukrainian-market-mobile-phones-data	31KB	2021-02-05 10:48:09	6792	59	1.0	Mobile Phones Data
pratikgarai/mobile-phone-specifications-and-prices	30KB	2022-08-14 23:25:36	3336	51	0.8235294	Mobile Phone Specifications and Prices
mfekadu/darpa-timit-acousticphonetic-continuous-speech	829MB	2019-06-05 02:01:06	9634	138	0.8235294	DARPA TIMIT Acoustic-Phonetic Continuous Speech
ahsan81/used-handheld-device-data	77KB	2023-01-07 16:47:53	2953	58	1.0	Used Phones & Tablets Pricing Dataset

msainani/gsmarena-mobile-devices					GSMarena Mobile Phone
Devices	1MB	2020-06-24 15:12:02		2681	56
0.9411765					
nkitgupta/flipkart-cell-phone-reviews					Flipkart Cell Phone
Reviews	3MB	2022-07-05 06:59:34		1160	
34 1.0					
devsubhash/flipkart-mobiles-dataset					Flipkart Mobiles Dataset
54KB	2022-03-20 10:46:28	13013	85	1.0	
ruchi798/mobile-phone-tweets					Mobile Phone Tweets
4MB	2021-12-15 14:15:02	502	34	1.0	

```
[9]: #downloading the data set
!kaggle datasets download -d "prasertk/mobile-phone-rating"
```

Dataset URL: <https://www.kaggle.com/datasets/prasertk/mobile-phone-rating>

```
0%|          | 0.00/3.35k [00:00<?, ?B/s]
100%|#####| 3.35k/3.35k [00:00<00:00, 757kB/s]
```

License(s): CC0-1.0

Downloading mobile-phone-rating.zip to

C:\Users\DONATUS\Desktop\EverythingDataScience\pythonML\fullcourse\part1

Unzip and Read the Data Unzip using zipfile library

```
[23]: import zipfile
```

```
[25]: with zipfile.ZipFile('mobile-phone-rating.zip', "r") as file:
        file.extractall('phones')
```

```
[26]: source_file_path = ""
        destination_path = ""
```

```
[29]: import os
        os.listdir("phones")
```

```
[29]: ['mobile phone rating by dxo.csv']
```

```
[30]: import pandas as pd
```

```
[32]: api= pd.read_csv('phones/mobile phone rating by dxo.csv')
```

```
[33]: api.head()
```

```
[33]:
```

	model	price	launch	camera	selfie	audio	display \
0	Huawei P50 Pro	\$907	Jul 2021	144.0	106.0	NaN	93.0
1	Xiaomi Mi 11 Ultra	\$1200	Mar 2021	143.0	94.0	71.0	87.0

2	Huawei Mate 40 Pro+	\$1363	Oct 2020	139.0	NaN	NaN	NaN
3	Apple iPhone 13 Pro Max	\$1099	Sep 2021	137.0	99.0	75.0	99.0
4	Apple iPhone 13 Pro	\$999	Sep 2021	137.0	99.0	75.0	98.0

	battery
0	NaN
1	69.0
2	NaN
3	89.0
4	76.0

EXAMPLE TWO FOLLOW THE STEPS

```
[35]: !kaggle datasets list -s "LANL Earthquake Prediction"
```

ref	size	lastUpdated	downloadCount	voteCount	usabilityRating	title
takeiy/data-for-lanl-earthquake-prediction	364MB	2019-05-26 11:52:25	171	7	0.23529412	DATA for LANL Earthquake Prediction
parushkin/lanl-lb	202KB	2019-06-11 06:49:25	22	1	0.3529412	LANL LB
bernir/bernisl-lanl-features	687MB	2019-05-27 06:35:45	36	2	0.75	Berni's LANL Features
tocha4/lanl-training-as-pickle	413MB	2019-04-13 09:12:13	104	1	0.25	LANL training as pickle

```
[ ]: #!kaggle datasets download -d "takeiy/data-for-lanl-earthquake-prediction"
```

```
[ ]: !kaggle competitions download -c LANL-Earthquake-Prediction
```

```
[ ]:
```