List of Popular Datasets Sites

1 Deep Learning / Machine Learning Datasets Sites

1. **UCI Machine Learning Repository**: A popular resource for datasets in the machine learning community.

   - https://archive.ics.uci.edu/ml/index.php

2. **Google Dataset Search**: A tool by Google that helps you find datasets stored across the web.

   - https://datasetsearch.research.google.com/

3. **AWS Public Datasets**: A collection of publicly available datasets hosted by Amazon Web Services.

   - https://registry.opendata.aws/

4. **Microsoft Azure Open Datasets**: A repository of open datasets curated by Microsoft.

   - https://azure.microsoft.com/en-us/services/open-datasets/

5. **OpenML**: An online platform for sharing and finding datasets, algorithms, and experiments.

   - https://www.openml.org/

6. **Data.gov**: The U.S. government's open data site, which includes datasets on various topics.

   - https://www.data.gov/

7. **Quandl**: A platform for financial, economic, and alternative datasets.

   - https://www.quandl.com/

8. **Kaggle Datasets**: Although you mentioned not to use Kaggle, it's worth noting for completeness.

   - https://www.kaggle.com/datasets

9. **Google Cloud Public Datasets**: A collection of public datasets hosted by Google Cloud.

   - https://cloud.google.com/public-datasets

10. **Papers with Code**: A site that links research papers with their associated datasets and code.

    - https://paperswithcode.com/datasets

11. **Machine Learning Datasets at Awesome Public Datasets**: A curated list of public datasets on GitHub.

    - https://github.com/awesomedata/awesome-public-datasets#machine-learning

12. **Stanford Large Network Dataset Collection (SNAP)**: Datasets for social network analysis and more.

    - https://snap.stanford.edu/data/

13. **KDNuggets Datasets**: A collection of datasets recommended by KDNuggets for machine learning projects.

    - https://www.kdnuggets.com/datasets/index.html

14. **AI2 Datasets by Allen Institute for AI**: A collection of datasets provided by the Allen Institute for AI.

    - https://allenai.org/data

2 Some IoT / Cyber Physical Systems Datasets for Machine Learning and Deep Learning

1. **UCI Machine Learning Repository - Smart Home Dataset**: Contains data from a real smart home setup including various sensors and environmental readings.

   - https://archive.ics.uci.edu/ml/datasets/Smart+Home+Dataset+with+Weather+Information

2. **Intel Berkeley Research Lab - Sensor Data**: Sensor data from a deployment of 54 sensors in the Intel Berkeley Research lab.

   - http://db.lcs.mit.edu/labdata/labdata.html

3. **Smart Data Set**: Data from various IoT devices deployed in real homes.

   - http://traces.cs.umass.edu/index.php/Smart/Smart

4. **IoT Network Intrusion Dataset (TON_IoT)**: A dataset for cybersecurity applications in IoT.

- https://research.unsw.edu.au/projects/toniot-datasets

5. **CASAS Smart Home Project**: Datasets from the CASAS smart home project, which includes sensor data for activity recognition.

    - http://casas.wsu.edu/datasets/

6. **CityPulse: Smart City Datasets**: A collection of datasets for various smart city applications, including transportation and environmental monitoring.

    - http://iot.ee.surrey.ac.uk:8080/datasets.html

7. **Pecan Street Dataport**: Smart grid data including residential energy consumption and solar power generation.

    - https://dataport.pecanstreet.org/

8. **Tennessee Eastman Process Simulation Data**: A dataset used for monitoring and fault diagnosis in industrial processes.

    - https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6C3JR1

9. **GreenButton Energy Data**: Standardized format for representing energy usage information.

    - https://www.greenbuttondata.org/

10. **Ocean Observatories Initiative (OOI) Data**: Oceanographic data collected from various sensors deployed in the ocean.

    - https://ooinet.oceanobservatories.org/

11. **Air Quality Data Set**: Real-time data on air quality parameters collected from various sensors.

    - https://archive.ics.uci.edu/ml/datasets/Air+Quality

12. **UK-DALE: UK Domestic Appliance-Level Electricity Dataset**: A dataset for understanding energy consumption patterns in households.

    - https://data.ukedc.rl.ac.uk/browse/edc/efficiency/residential/Electricity-Use

13. **IEEE DataPort: Datasets for IoT and Cyber-Physical Systems**: A collection of datasets for various IoT and CPS applications.

    - https://ieee-dataport.org/

14. **OpenSky Network - Air Traffic Data**: Air traffic surveillance data for monitoring and analysis.

- https://opensky-network.org/datasets

15. **SECOM Manufacturing Data**: Dataset used for anomaly detection in semiconductor manufacturing.

    - https://archive.ics.uci.edu/ml/datasets/SECOM

16. **NASA Prognostics Data Repository**: Datasets for the study of prognostics and health management.

    - https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository

17. **Smart Cities Open Data Portal**: A portal offering various datasets related to smart cities applications.

    - https://data.smartcity.barcelona/

18. **Aranet4: Environmental Sensor Data**: Real-time environmental data including temperature, humidity, and CO2 levels.

    - https://aranet.com/en/download-category/environmental-data/

19. **MeteoNet: Weather Prediction Data**: A dataset for weather prediction using IoT sensor data.

    - https://meteonet.umr-cnrm.fr/dataset.html

20. **OpenAIRE: Research Data for IoT and CPS**: A repository of open access research data for IoT and CPS applications.

    - https://www.openaire.eu/