

Day 3:

Working with Real Data:

In "Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron, "Working with Real Data" emphasizes the practical aspect of applying machine learning techniques to real-world datasets. This section guides readers through the essential steps of data preprocessing, including cleaning, handling missing values, scaling, and encoding categorical features. It also covers strategies for splitting data into training and testing sets, along with techniques for cross-validation. Géron underscores the importance of understanding data thoroughly to build reliable and effective machine learning models, providing readers with the foundational knowledge needed to tackle real-world problems confidently.

- Popular open data repositories:

- UC Irvine Machine Learning Repository

- Kaggle datasets

- Amazon's AWS datasets

- Meta portals (they list open data repositories):

- <http://dataportals.org/>

- <http://opendatamonitor.eu/>

- <http://quandl.com/>

- Other pages listing many popular open data repositories:

- Wikipedia's list of Machine Learning datasets

- Quora.com question

- Datasets subreddit