



Published on *Java Machine Learning Library (Java-ML)* (<http://java-ml.sourceforge.net>)

[Home](#) > Databases

By *Thomas Abeel*

Created 12/17/2008 - 12:13

Databases

We provide mirrors for a number of well known datasets and we host a number of datasets that have been used in the scientific literature for validation.

UCI datasets

One of the most well-known repositories of machine learning related datasets is the [UCI Machine Learning Repository](#) ^[1]. Currently they host over 170 datasets related to number of machine learning fields including classification, clustering and regression.

We provide two packages, first a collection of 111 'small' datasets that contain less than 10 Mb of data, and second a set of 7 larger datasets which have over 10 Mb of data.

[Download 111 small UCI datasets](#) ^[2]

[Download 7 large UCI datasets](#) ^[3]

For full details on all datasets we would like to refer to the [home page](#) of the UCI repository. If you use these datasets in your research, please [cite the original authors work](#) ^[1].

Copyright 2006-2009 [Thomas Abeel](#)

Source URL (retrieved on 08/20/2009 - 17:54): <http://java-ml.sourceforge.net/content/databases>

Links:

[1] <http://mlr.cs.umass.edu/ml/>

[2] <http://downloads.sourceforge.net/java-ml/UCI-small.zip>

[3] <http://downloads.sourceforge.net/java-ml/UCI-large.zip>