Machine Data Hub

Release June 2021

Cecilia Barnes, Peter Sulcs

May 27, 2021

CONTENTS:

ABOUT

The Machine Data Hub python library is an open source project for viewing, downloading, and suggesting datasets for Prognostics Health Management Machine Learning Projects.

Eventually, the goal is to have machine-data-hub complete ETL processes on datasets to allow users to download clean, easy to use data.

TWO

INSTALLATION

To install:

```
pip install machine-data-hub
```

GUIDE

3.1 Tutorial

Basic Commands to use machine-data-hub below.

3.1.1 To Begin

First, install the library:

```
pip install machine-data-hub
```

3.1.2 Commands

View All Datasets

Machine Data Hub allows you to view a library of reliable and useful machine related datasets. To view all datasets:

```
$ mdh list
```

This will list out the dataset ID, name, number of individual files, and file size for each dataset.

Get Metadata

If you want to know more about a dataset, you can view metadata such as ML Type, if it's a time series, if it's labeled, and much more. To view a dataset's metadata:

```
$ mdh metadata <id>
```

Example to view metadata for dataset with ID of 2:

```
$ mdh metadata 2
```

Download a Dataset

If you find a dataset that you like and want to use for your project, you can download it. To download a full dataset:

```
$ mdh download <id>
```

If the dataset has multiple files, you can just download one of the files at once. To download one specific file in a dataset:

```
$ mdh download <id> <file number>
```

Example to download file 1 from dataset 15:

```
$ mdh download 15 1
```

Suggest a Dataset

Do you know a great dataset that could be useful for prognostics or other machine related projects? Suggest it be added to the Machine Data Hub's library of datasets:

```
$ mdh suggest <link> <name> <summary>
```

Example to download file 1 from dataset 15:

```
$ mdh suggest "www.data.com" "Dataset" "This is a summary"
```

3.2 Release Notes

3.2.1 Release 0.1.0

Release Date:

What's New

- First commit to publish to PyPI, reserving package name
- Functionality to view and download datasets from the Machine Data Hub.

3.3 License

MIT License

Copyright (c) 2020 PHM-Data-Hub

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

FOUR

INDICES AND TABLES

- genindex
- modindex
- search