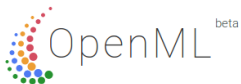


Heidi Seibold

Lunch time lecture, October 2017



Exploring machine learning better, together

19633  
data sets

Find or add **data** to analyse

25946  
tasks

Download or create scientific  
**tasks**

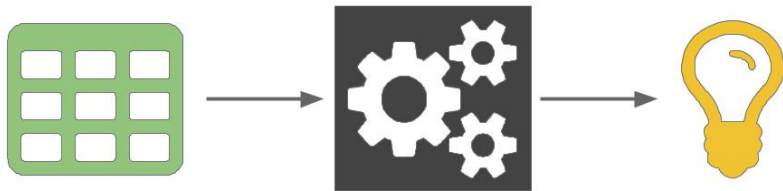
2991  
flows

Find or add data analysis **flows**

1705710  
runs

Upload and explore all **results**  
online.

# Machine Learning



Data analysis and machine learning are important in research fields where data are collected.

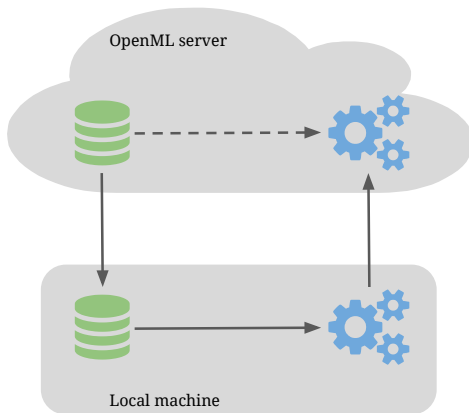


But many researchers need help with this.

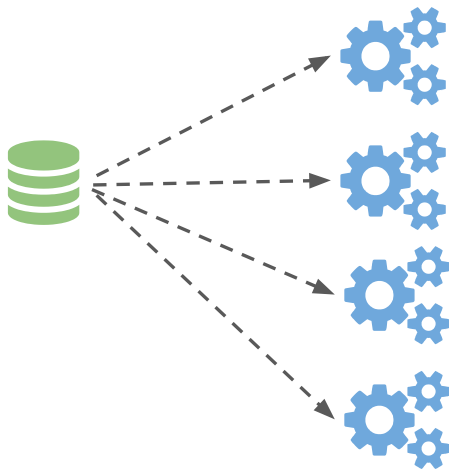


OpenML can be a way to ask machine learners all over the world for help.

# OpenML machine learner workflow

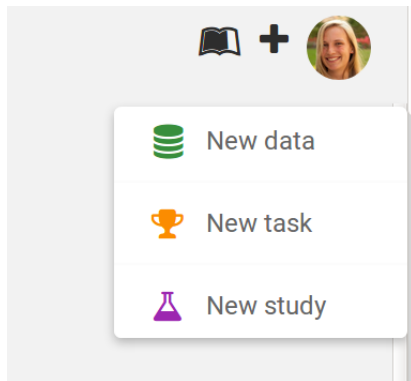


# One question, many answers



⇒ pick the best!

# Upload a data set





# Upload a data set



## Add data

### Data files

Browse...

Or URL (not both)



**Name**

**Version**

### Description

Short description (can still be edited online).  
Use #tags to label it. Describe where the  
data originates from, and whether it was  
processed in any way.

**Data format** ([see specifications](#))

ARFF

### Author(s)

**Licence - [Learn more](#)**

Publicly available

Mark a work that is free of known copyright  
restrictions. [More info](#)

### Citation requests

How to reference this data in future work  
(e.g., publication, DOI).

### Who can view this data

**Under development**

Everyone

# Create a task

task = the question = prediction problem

0 tasks

DEFINE A NEW TASK

# Create a task



## Create new task

### Choose Task Type

#### Task types

Supervised Classification

In supervised classification, you are given an input dataset in which instances are labeled with a certain class. The goal is to build a model that predicts the class for future unlabeled instances. The model is evaluated using a train-test procedure, e.g. cross-validation.

### Task inputs

#### Dataset(s)

Run\_or\_walk\_information(1)



#### Target Feature

activity

#### Estimation Procedure









10 TIMES 10-FOLD CROSSVALIDATION ▾

Now is your turn. Try OpenML and let us know what you think!

<https://github.com/openml/OpenML/issues>

# Want to get involved?

Start by participating in hacktoberfest!

<input type="checkbox"/>	 <b>Upload your favorite data set to OpenML and let us know if something does not work</b> <b>beginner</b> <b>Hacktoberfest</b> #461 opened 19 minutes ago by HeidiSeibold
<input type="checkbox"/>	 <b>Guide: can my data be used for machine learning?</b> <b>Hacktoberfest</b> #459 opened a day ago by HeidiSeibold
<input type="checkbox"/>	 <b>Pathway for a person who has data and a question and wants to use OpenML to get an answer</b> <b>Hacktoberfest</b> #458 opened a day ago by HeidiSeibold
<input type="checkbox"/>	 <b>Create tool to convert different data types to ARFF</b> <b>Hacktoberfest</b> #457 opened a day ago by HeidiSeibold
<input type="checkbox"/>	 <b>Explain how to prepare data for upload to OpenML (convert different data types to ARFF)</b> <b>document it</b> <b>Hacktoberfest</b> #456 opened a day ago by HeidiSeibold
<input type="checkbox"/>	 <b>Making OpenML understandable for people with data (who aren't machine learners)</b> <b>Hacktoberfest</b> #455 opened a day ago by HeidiSeibold
<input type="checkbox"/>	 <b>Improve user-friendliness of openml.org</b> <b>beginner</b> <b>Hacktoberfest</b> #454 opened a day ago by HeidiSeibold
<input type="checkbox"/>	 <b>Improve OpenML guide</b> <b>beginner</b> <b>Hacktoberfest</b> #453 opened a day ago by HeidiSeibold

<https://github.com/openml/OpenML/issues>

# Questions? Comments?

Contact:

Heidi: [heidi.seibold@uzh.ch](mailto:heidi.seibold@uzh.ch)  
@HeidiBaya

OpenML: [openmlhq@googlegroups.com](mailto:openmlhq@googlegroups.com)  
@open\_ml