



OpenML

DEMOCRATIZING AND AUTOMATING
MACHINE LEARNING

Research different.

Polymaths: Solve math problems
by massive **online** collaboration

Broadcast question, combine
many minds to solve it

Networked Science

Serendipity: what's hard for one person is easy for another
Collaboration only scales if **all friction is eliminated**

Easy, organized, access to data, code, and results



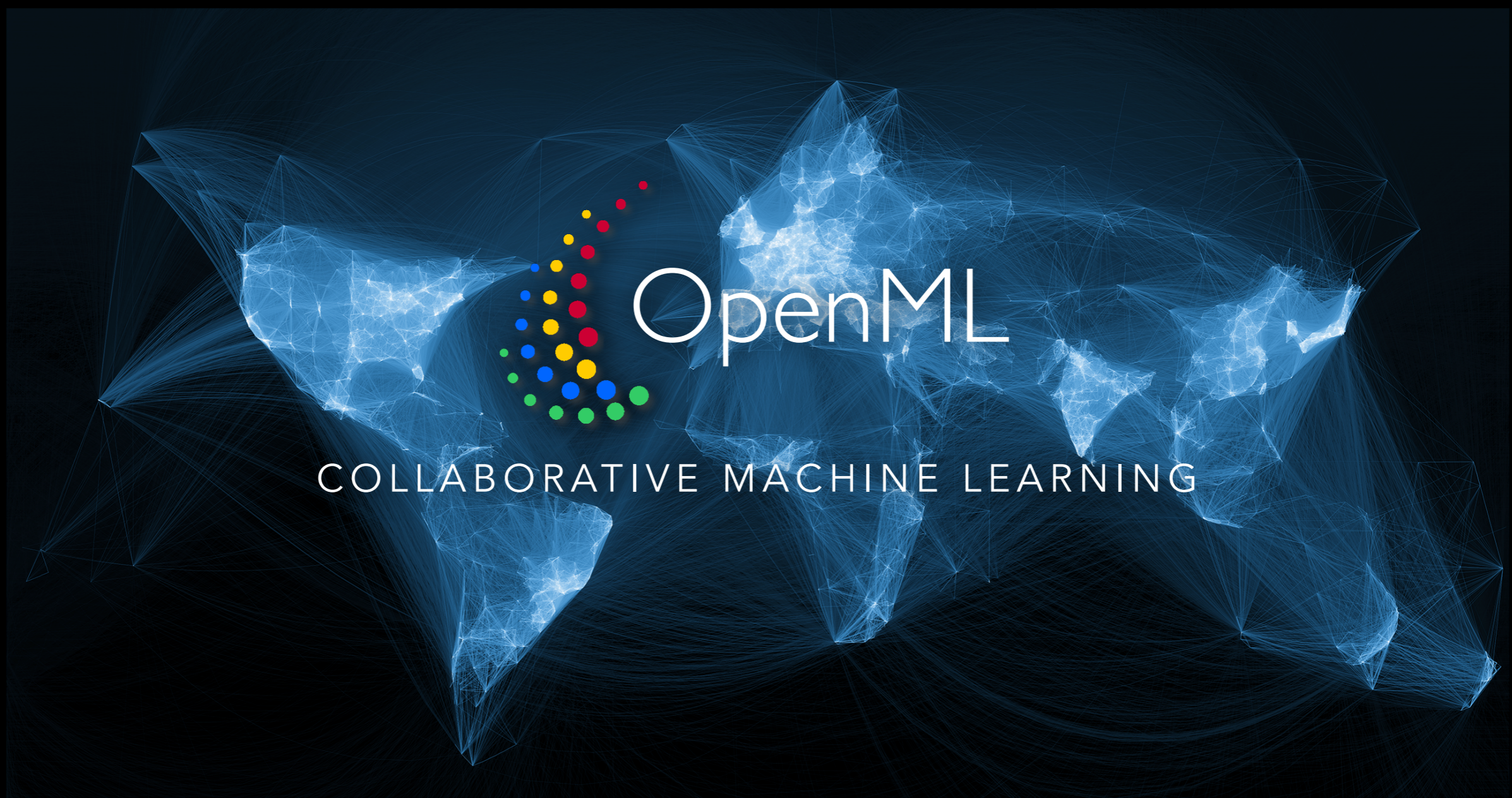
WHAT IF WE CAN EXPLORE DATA
COLLABORATIVELY



WHAT IF WE CAN EXPLORE DATA
COLLABORATIVELY
ON WEB SCALE



WHAT IF WE CAN EXPLORE DATA
COLLABORATIVELY
ON WEB SCALE IN REAL TIME



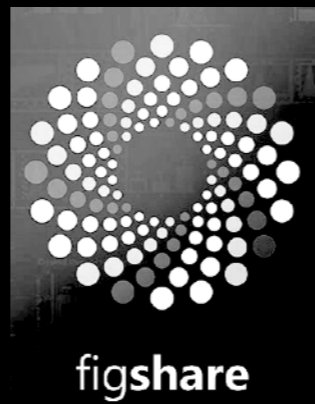
Easy to use: Integrated in many ML environments

Easy to contribute: Automated sharing of data, code, results

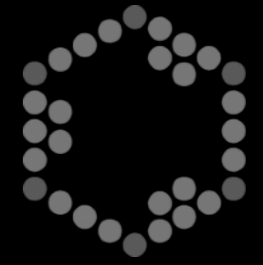
Organized data: Reproducible, connected to data, code, people

Reward structure: Build reputation and trust

Self-learning: Learn from millions of experiments to help users



zenodo



mldata.org
machine learning data set repository

kaggle



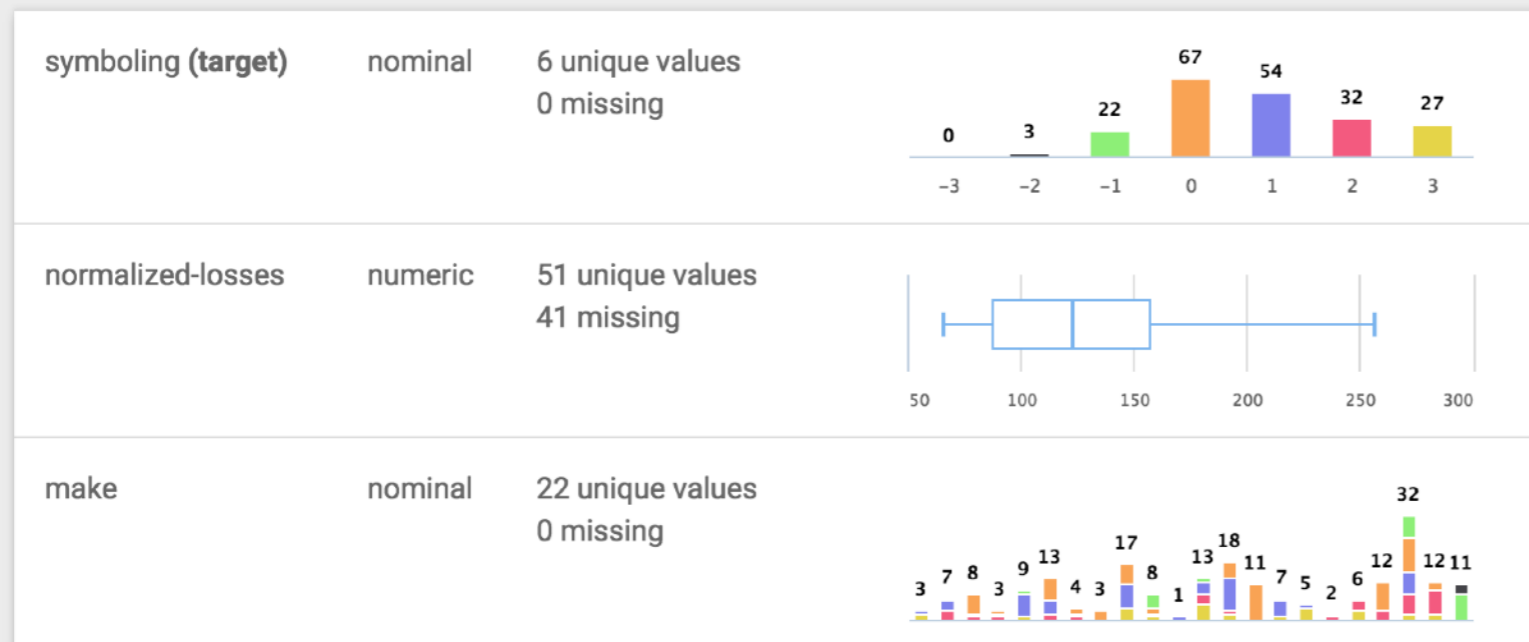
Data (ARFF) uploaded or referenced, versioned
analyzed, characterized, organized online



analyzed, characterized, organized online

+ visualizations, statistics, landmarks, error checking,
queryable through website + API

26 features



▼ Show all 26 features

72 properties

DefaultAccuracy	0.33	The predictive a
NumberOfClasses	7	The number of c
NumberOfFeatures	26	The number of f
NumberOfInstances	205	The number of i
NumberOfMissingVal...	59	Counts the total



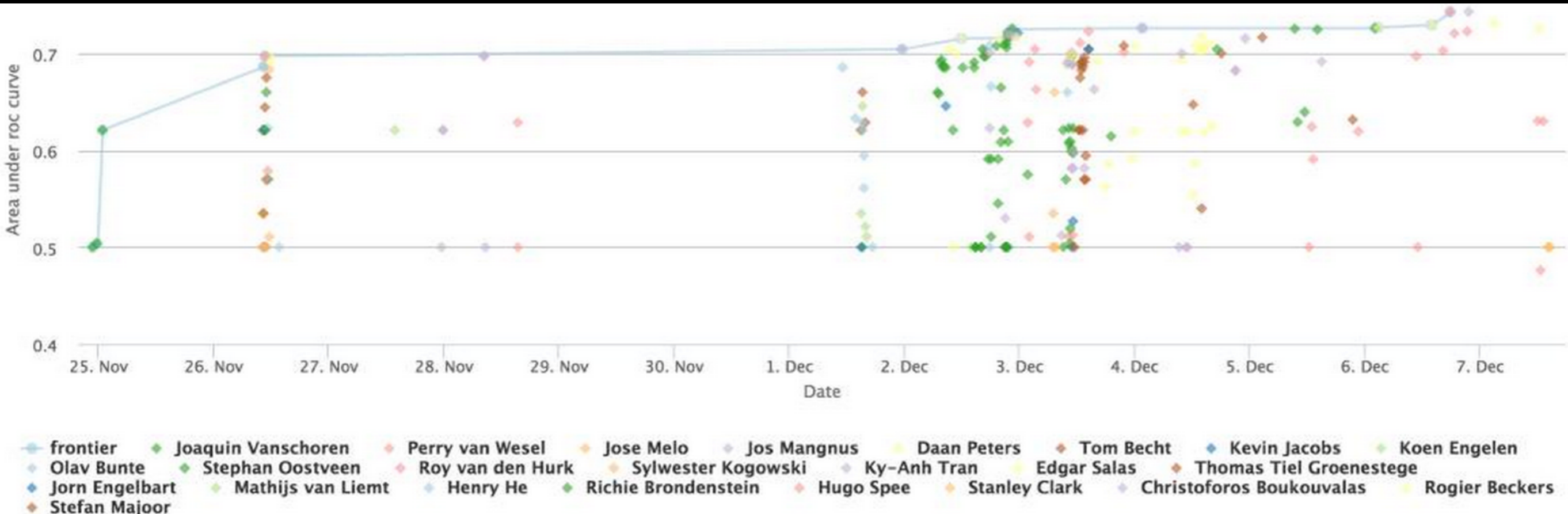
Tasks contain data, goals, procedures.

Readable by tools, automates experimentation

All results organized online: **realtime overview**



All results organized online: **realtime overview**

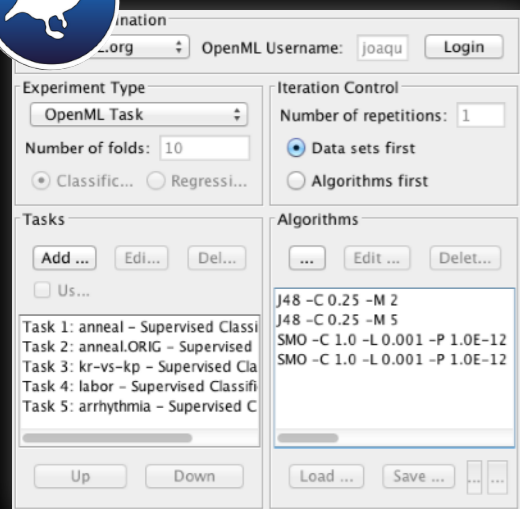




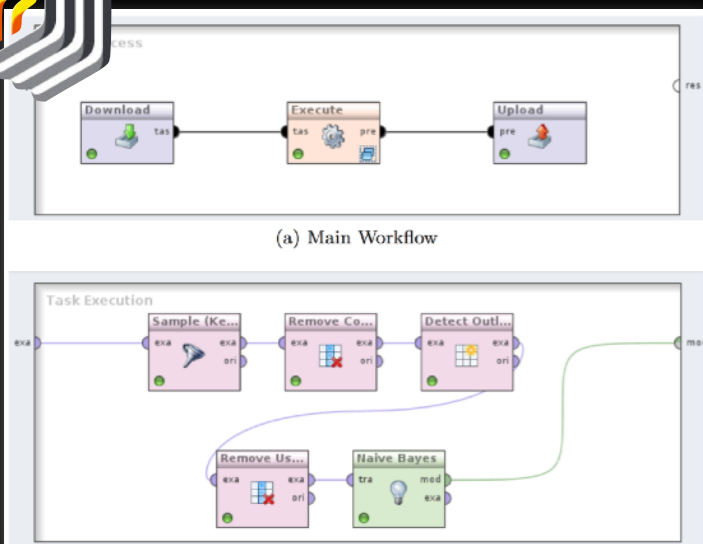
Flows (code) run anywhere, using your favorite tools
Integrations + APIs (REST, R, Python, Java,...)



Integrations + APIs (REST, R, Python, Java,...)



```
from sklearn import tree
from openml import tasks, runs
task = tasks.get_task(14951)
clf = tree.DecisionTreeClassifier()
run = runs.run_task(task, clf)
return_code, response = run.publish()
```



```
library(OpenML)
library(mlr)
task = getOMLTask(10)
lrn = makeLearner("classif.rpart")
res = runTaskMlr(task, lrn)
run.id = uploadOMLRun(res)
```




Integrations + APIs (REST, R, Python, Java,...)



```
from sklearn import tree
from openml import tasks,runs
task = tasks.get_task(14951)
clf = tree.DecisionTreeClassifier()
run = runs.run_task(task, clf)
return_code, response = run.publish()
```





Experiments auto-uploaded, evaluated online
reproducible, linked to **data**, **flows**, **authors**
and **all other experiments**



Experiments auto-uploaded, evaluated online

Result files



Description

XML file describing the run, including user-defined evaluation measures.



Model readable

A human-readable description of the model that was built.



Model serialized

A serialized description of the model that can be read by the tool that generated it.



Predictions

ARFF file with instance-level predictions generated by the model.

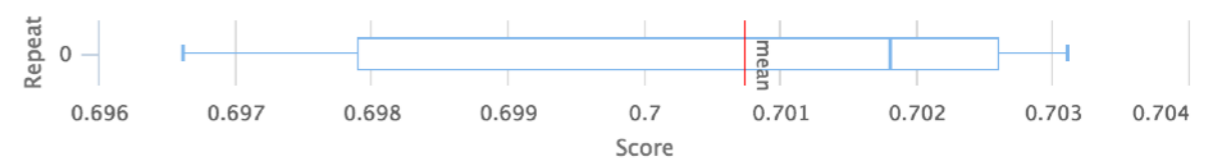
Area under ROC curve

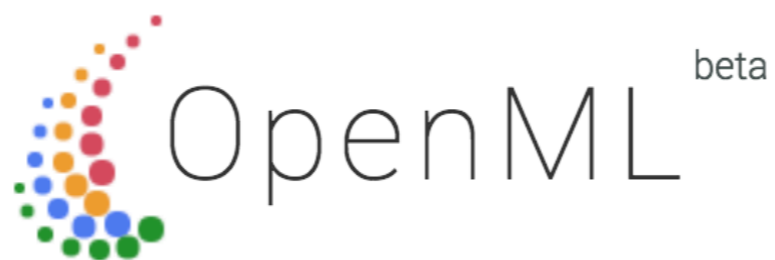
0.7007 \pm 0.0023

Per class

0	1
0.7007	0.7007

Cross-validation details (10-fold Crossvalidation)





Exploring machine learning better, together

19626
data sets

Find or add **data** to analyse

25941
tasks

Download or create scientific **tasks**

3353
flows

Find or add data analysis **flows**

1698629
runs

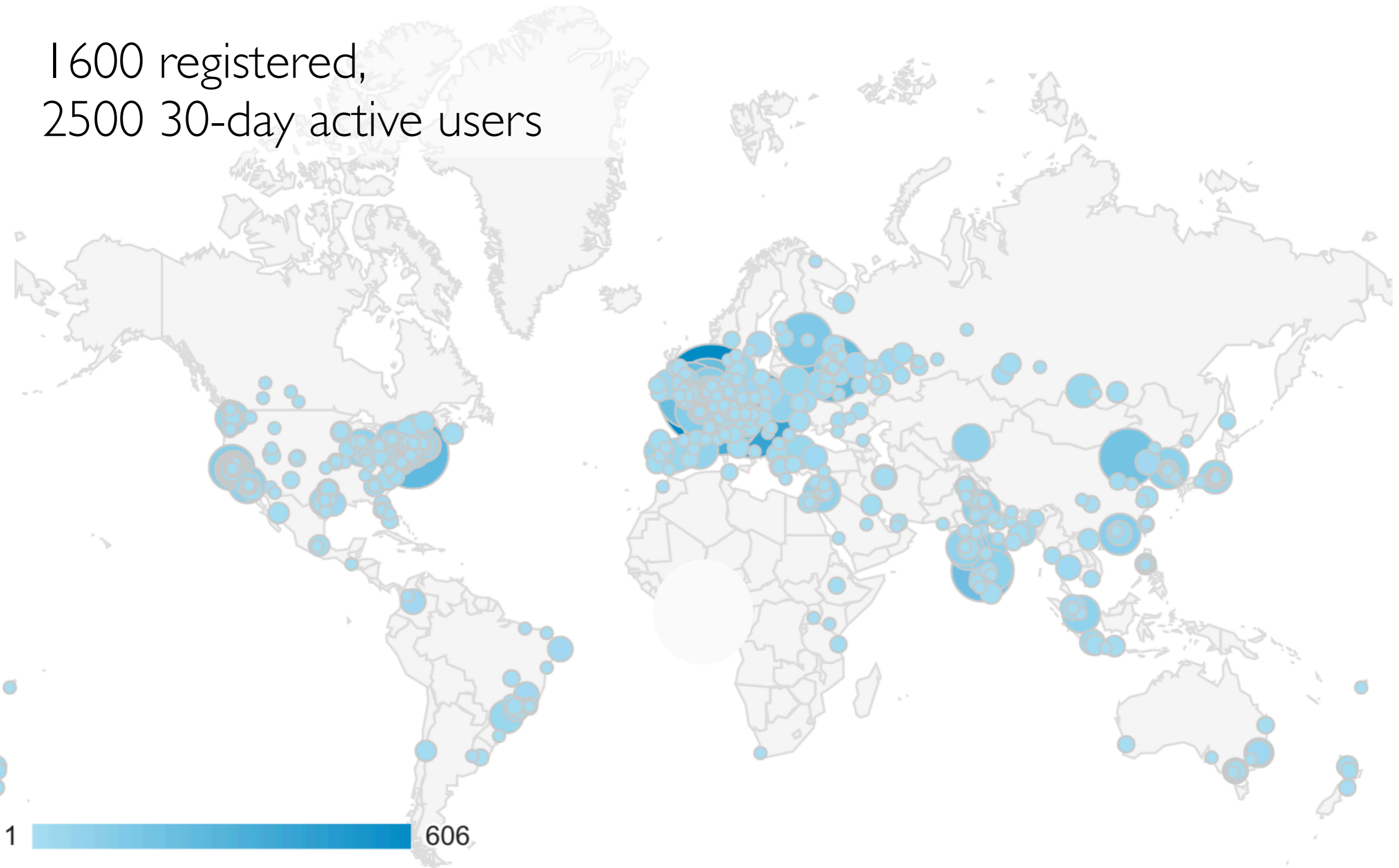
Upload and explore all **results** online.

Download and share data, flows and runs through:



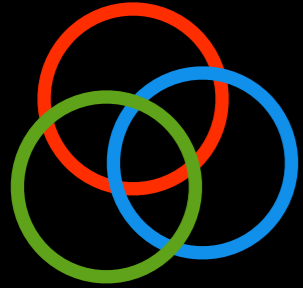
OpenML Community

1600 registered,
2500 30-day active users



Jul-Nov 2016

Collaboration tools (in progress)



Circles

Create collaborations with trusted researchers



Studies (e-papers)

Online counterpart of a paper, linkable



Reputation

Auto-tracking of your activity, reach, impact



Notebooks

Easy online collaboration on data analysis scripts

Join Us!

www.openml.org

Join our hackathons

 @open_ml

 OpenML



Thank You

