

You can have fun doing  
Kaggle competitions.

Let's Start

pedro.lima@cognitiva.com



# cognitiva



[www.cognitiva.com](http://www.cognitiva.com)

kaggle Search Competitions Datasets Kernels Discussion Learn ... Sign In

## Competitions

Documentation InClass

General InClass Sort by Grouped All Categories Search competitions

### 20 Active Competitions

 TWO SIGMA	<b>Two Sigma: Using News to Predict Stock Movements</b> Use news analytics to predict stock price performance <small>Featured · Kernels Competition · 2 months to go · news agencies, time series, finance, money</small>	\$100,000 2,927 teams
	<b>Jigsaw Unintended Bias in Toxicity Classification</b> Detect toxicity across a diverse range of conversations <small>Featured · Kernels Competition · a month to go · biases, nlp, text data</small>	\$65,000 2,242 teams
	<b>LANL Earthquake Prediction</b> Can you predict upcoming laboratory earthquakes? <small>Research · 9 days to go · earth sciences, physics, signal processing</small>	\$50,000 4,297 teams
	<b>Google Landmark Recognition 2019</b> Label famous (and not-so-famous) landmarks in images <small>Research · 9 days to go</small>	\$25,000 243 teams

# Agenda

Introduction to Kaggle

Kaggle workshop

kaggle Search  Competitions Datasets Kernels Discussion Learn ... Sign In

## House Prices: Advanced Regression Techniques



Predict sales prices and practice feature engineering, RFs, and gradient boosting  
4,553 teams · Ongoing

Overview Data Kernels Discussion Leaderboard Rules [Join Competition](#)

### Overview

**Description**

**Evaluation**

**Tutorials**

**Frequently Asked Questions**

**Start here if...**

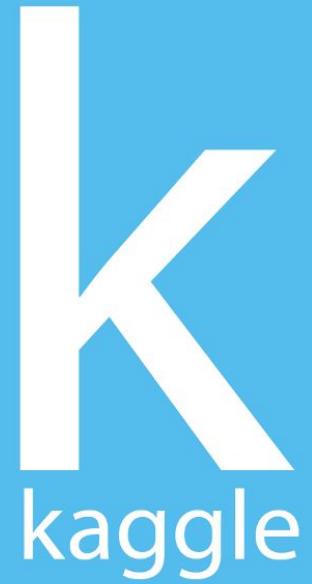
You have some experience with R or Python and machine learning basics. This is a perfect competition for data science students who have completed an online course in machine learning and are looking to expand their skill set before trying a featured competition.

**Competition Description**



Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.



It's ok to cut some corners

# Kaggle Platform

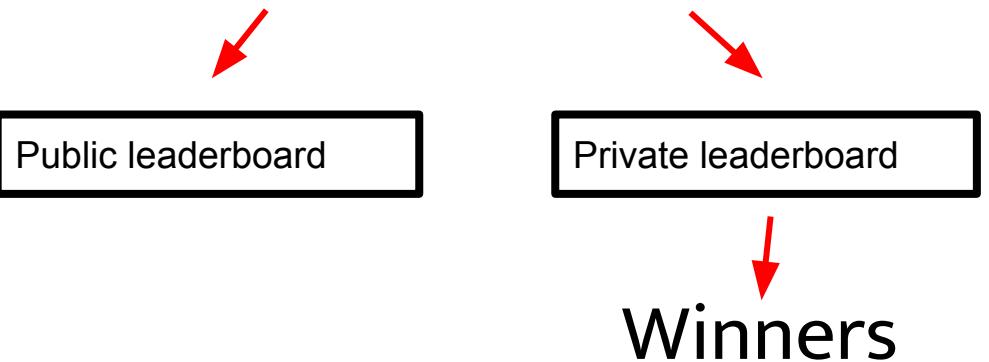
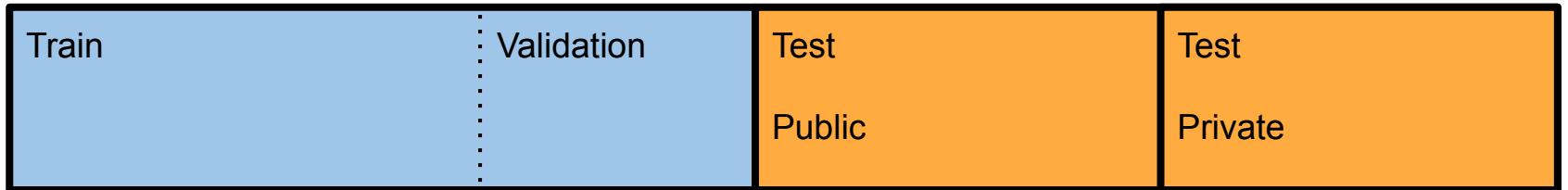
The image displays three screenshots of the Kaggle platform, illustrating its features for data science competitions.

**Top Left Screenshot:** Shows the "Toxic Comment Classification Challenge" page. It features a banner with a prize of \$35,000. Below the banner, there's a "Step 1: Upload submission file" section with an "Upload Files" button and instructions for CSV format. A "Step 2: Describe submission" section is partially visible at the bottom.

**Bottom Center Screenshot:** Shows the "Exploring running events data" tool. It includes a sidebar with "Datasets", "Workspaces", "Variables", and "Settings". The main area displays a list of event names such as "Facebook like", "Twitter like", "Twitter retweet", etc., with some code snippets below.

**Top Right Screenshot:** Shows the "Jigsaw Unintended Bias in Toxicity Classification" competition page. It features a banner with a prize of \$65,000. The "Leaderboard" tab is selected, showing a table of team scores. The table includes columns for Team Name, Kernel, Team Members, Score #, Entries, and Last. The top entries are:

Team Name	Kernel	Team Members	Score #	Entries	Last
pyrali			0.94540	141	10
FFRAP			0.94497	175	10
Kazhira			0.94445	30	49
[redacted] anyone tried best??			0.94444	104	50
Abhilashk			0.94392	58	38m
LimerickEDU@EDU			0.94327	2	42
We Need SGD-V100			0.94294	128	50
Tam Astrow			0.94282	45	10d
SQL & WSL			0.94279	123	11



# There is life outside Kaggle

codalab.org

drivendata.org

crowdanalytix.com

signate.jp (Japan)

topcoder.com

boosters.pro (Russia)

hackerearth.com

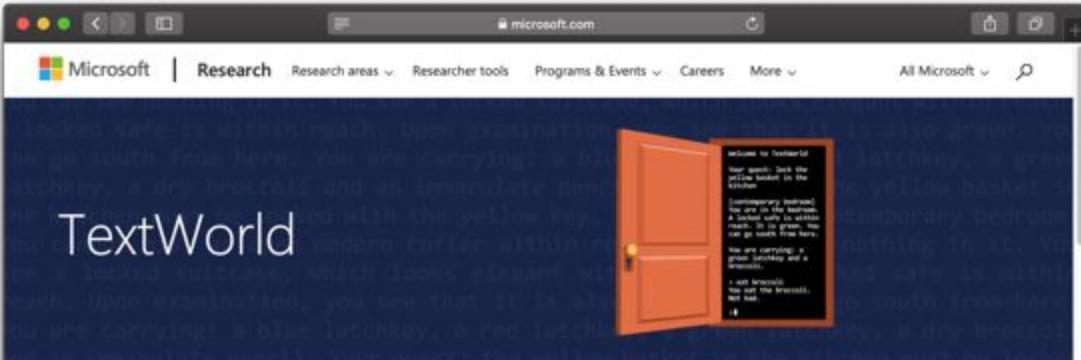
crowdai.org (Switzerland)

challenger.ai (China)

unearthed.solutions (Australia)

mafatchallenge.mod.gov.il (Israel)

(codalab is open source)



[Overview](#) [People](#) [Publications](#) [Microsoft Research blog](#) [In the news](#) [Contribute](#) [Try it](#)

Microsoft TextWorld is an open-source, extensible engine that both generates and simulates text games. You can use it to train reinforcement learning (RL) agents to learn skills such as language understanding and grounding, combined with sequential decision making.

You are navigating through a house. You've just entered a serious study. There is a gross looking mantle in the room. It has nothing on it. You see a closed rusty toolbox. Now why would someone leave that there?

Looks like there is a locked door. Find the key to unlock the door. You should try going east.

```
train — git-glulx-ml · tw-play tw-cooking-recipe1+go12-mmKotpb5foqKUvqm.ulx — 80x41
```

You are hungry! Let's cook a delicious meal. Check the cookbook in the kitchen for the recipe. Once done, enjoy your meal!

= Corridor =  
You've entered a corridor.

There is an exit to the east. Don't worry, there is no door. There is an exit to the south. Don't worry, there is no door. There is an exit to the west.

[> go south]

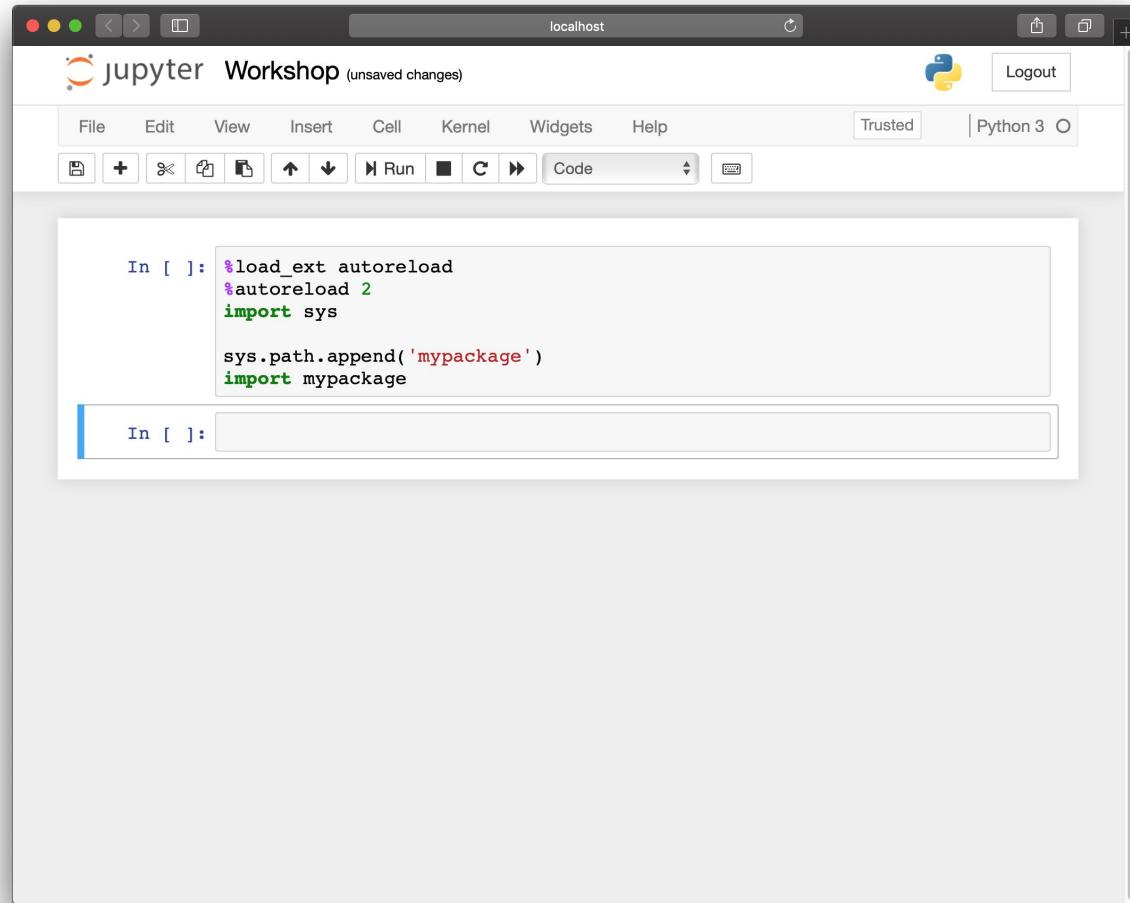
= Kitchen =  
You are in a kitchen. A typical one. You start to take note of what's in the room.

What's that over there? It looks like it's a fridge. What a letdown! The fridge is empty! You make out an oven. You make out a table. The table is massive. But there isn't a thing on it. You can make out a counter. The counter is vast. On the counter you can see a cookbook. You shudder, but continue examining the room. You rest your hand against a wall, but you miss the wall and fall onto a stove. The stove is conventional. But oh no! there's nothing on this piece of garbage. This always happens, here in TextWorld!

There is an open frosted-glass door leading east. There is an open patio door leading south. You need an exit without a door? You should try going north. You don't like doors? Why not try going west, that entranceway is not blocked by one.

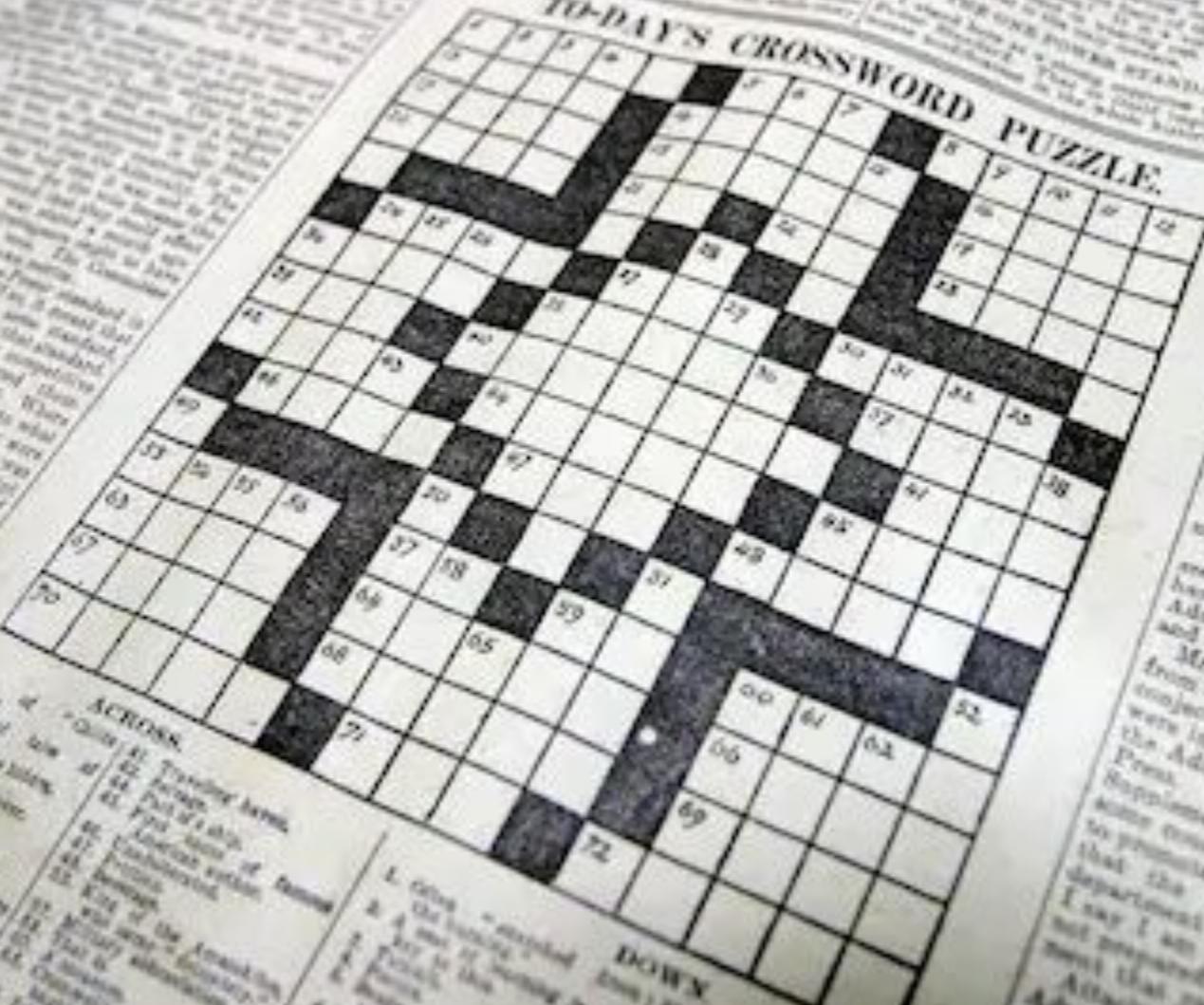
# Tools & Concepts

Scikit Learn / Xgboost / LightGBM / PyTorch  
K-Fold cross-validation  
Out-of-fold predictions  
Model Stacking  
Transfer learning / Pre-trained models



Don't take it too seriously





**ACROSS**

ACROSS

- |    |  |
|----|--|
| 42 | Transitory Areas                             |
| 43 | Parishes                                     |
| 44 | Part of which                                |
| 45 | Parishioners                                 |
| 46 | Parishioners outside of Parochial boundaries |
| 47 | Parishioners outside of Parochial boundaries |
| 48 | Parishioners                                 |
| 49 | Parishioners                                 |
| 50 | Parishioners                                 |
| 51 | Parishioners                                 |
| 52 | Parishioners                                 |
| 53 | Parishioners                                 |
| 54 | Parishioners                                 |
| 55 | Parishioners                                 |
| 56 | Parishioners                                 |

**DOWN**

Higgs Boson Machine Learning Challenge

Use the ATLAS experiment to identify the Higgs boson  
\$13,000 · 1,785 teams · 5 years ago

Overview Data Kernels Discussion Leaderboard Rules Join Competition

Public Leaderboard Private Leaderboard

The private leaderboard is calculated with approximately 82% of the test data.

This competition has completed. This leaderboard reflects the final standings.

Refresh

In the money Gold Silver Bronze

#	△pub	Team Name	Kernel	Team Members	Score	Entries	Last
1	▲ 1	Gábor Melis			3.80581	110	5y
2	▲ 1	Tim Salimans			3.78912	57	5y
3	▲ 1	nhlx5haze			3.78682	254	5y
4	▲ 38	ChoKo Team			3.77526	216	5y
5	▲ 35	cheng chen			3.77383	21	5y
6	▲ 16	quantify			3.77086	8	5y
7	▲ 1	Stanislav Semenov & Co (HS...)			3.76211	68	5y
8	▼ 7	Luboš Motl's team			3.76050	589	5y

 Higgs Boson Machine Learning Challenge

Use the ATLAS experiment to identify the Higgs boson  
\$13,000 · 1,785 teams · 5 years ago

Overview Data Kernels Discussion Leaderboard Rules New Topic



Bing Xu  
69th place

Hi all,

Tianqi Chen (crowswork) has made a fast and friendly boosting tree library [XGBoost](#). By using XGBoost and run a script, you can train a model with 3.60 AMS score in about 42 seconds.

The demo is at: <https://github.com/tqchen/xgboost/tree/master/demo/kaggle-higgs>, you can just type ./run.sh to get the score after you build it.

XGBoost is as easy to use as scikit-learn. And on my computer with Core i5-4670K CPU, the speed test.py (boosting 10 trees) shows:

```
sklearn.GBM costs: 77.5 seconds
XGBoost with 1 thread costs: 11.0 seconds
XGBoost with 2 thread costs: 5.85 seconds
XGBoost with 4 thread costs: 3.40 seconds
```

Like competitions held before, public sharing method will boost the performance of all teams and reduce barriers for new learners. We hope all of us can learn and enjoy more during the competition.

BTW, Don't forget to star [XGBoost](#) ;)

Update:

20th, May, 2014: If you are using XGBoost 0.2, please pull the newest version. The binary classification

*dmlc*  
**XGBoost**

K Keras