

HOW TO WIN A KAGGLE COMPETITION



ChrisCC

Calgary, AB, Canada

Joined 2 years ago · last seen in the past day



Followers 4



Competitions
Master

[Home](#)[Competitions \(14\)](#)[Kernels \(31\)](#)[Discussion \(31\)](#)[Followers \(4\)](#)[Contact User](#)[Follow User](#)

Competitions Summary



Competitions
Master

Current Rank

515

of 64,847

Highest Rank

309



Competitions: 14
Solo: 14 (100%)
Team: 0

Data Scientist@Shaw
Kaggle Master

KAGGLE?

Your Home for Data Science

Kaggle helps you learn, work, and play



Create an account

or

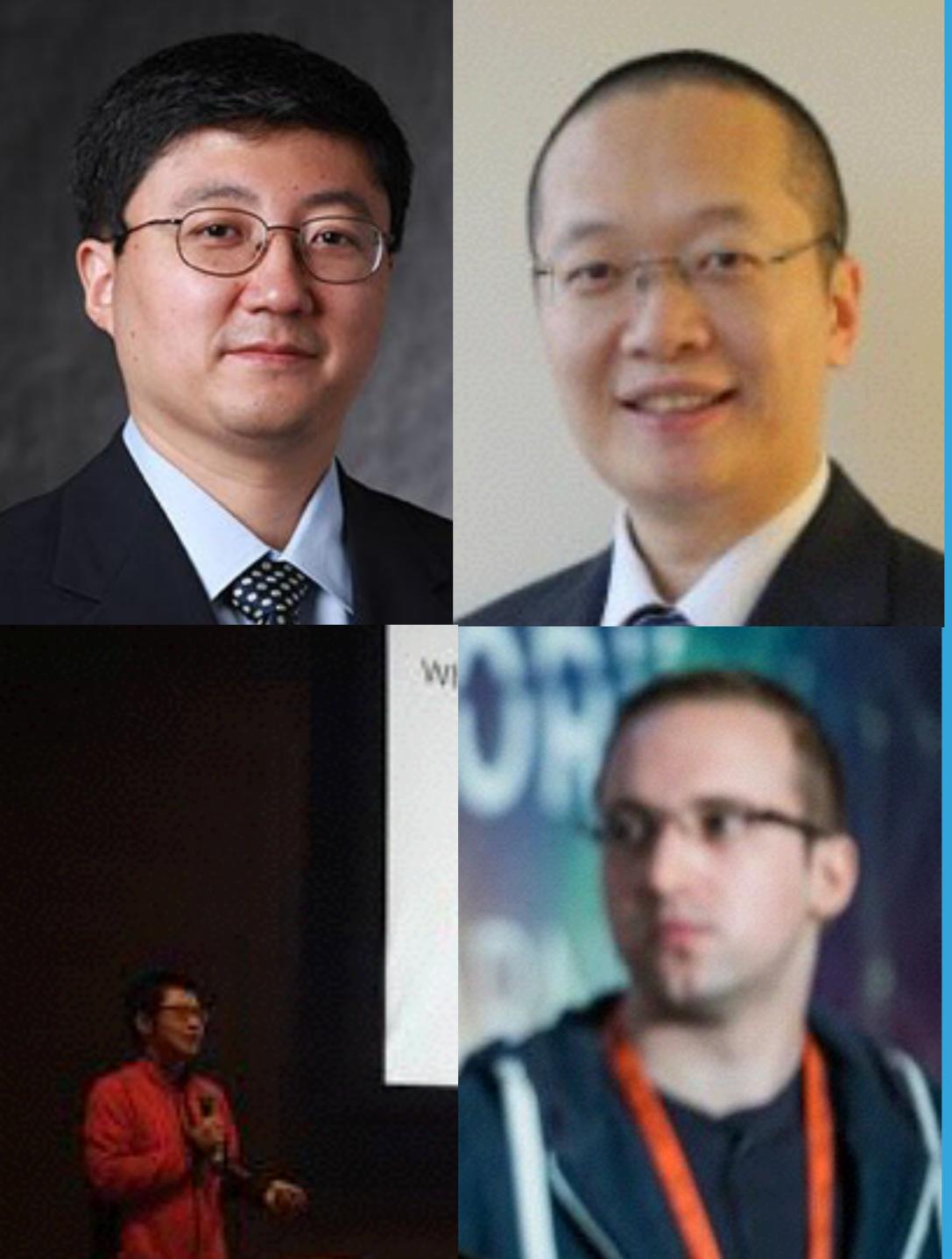
Host a competition

Largest and most influencing data science community

Battlefield for talented data scientists from 190 countries

Cradle of pioneering algorithms and approaches

WHY?



LEARN FROM THE BEST

Tong Zhang

#1 in "Heritage Health Prize"

Professor of Rutgers University

Writer of "Regularized Greedy Forest"

Tianqi Chen

#1 in "KDD Cup 2012"

#2 in "National Data Science Bowl"

Writer of XGBoost

Owen Zhang

Former #1 Kaggle

CPO @DataRobot

Jeffrey De Fauw

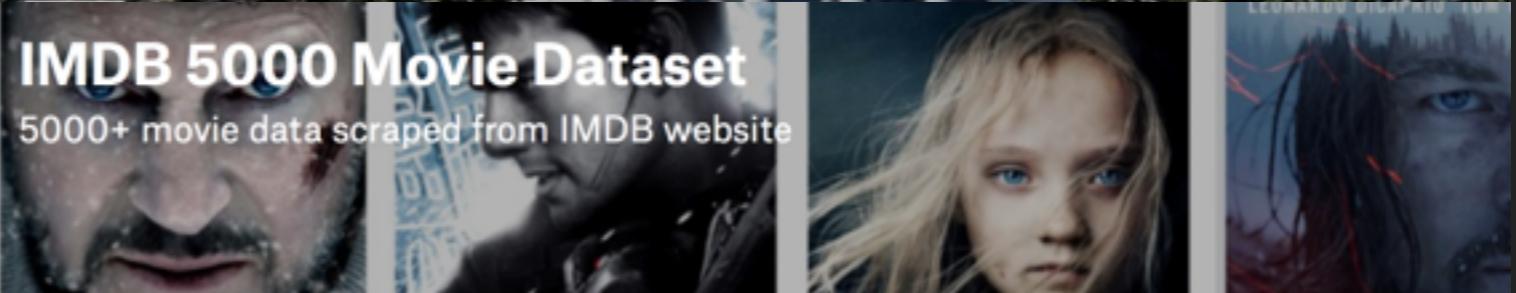
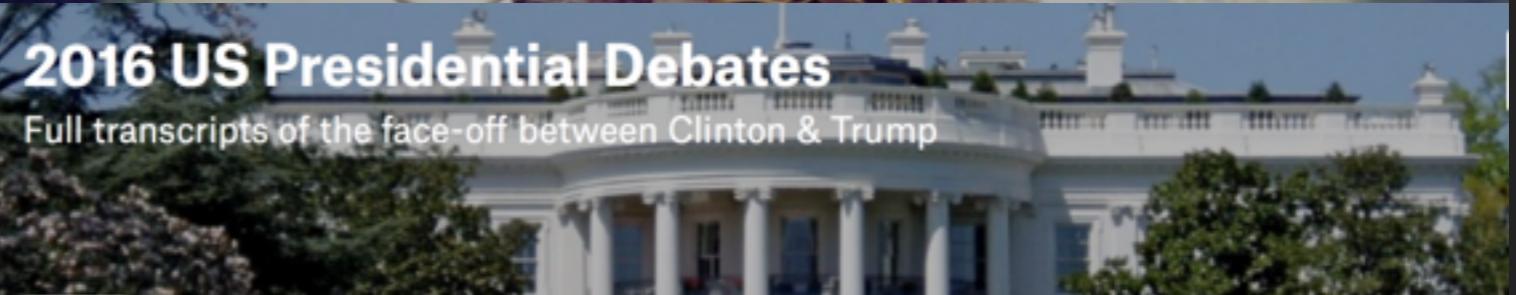
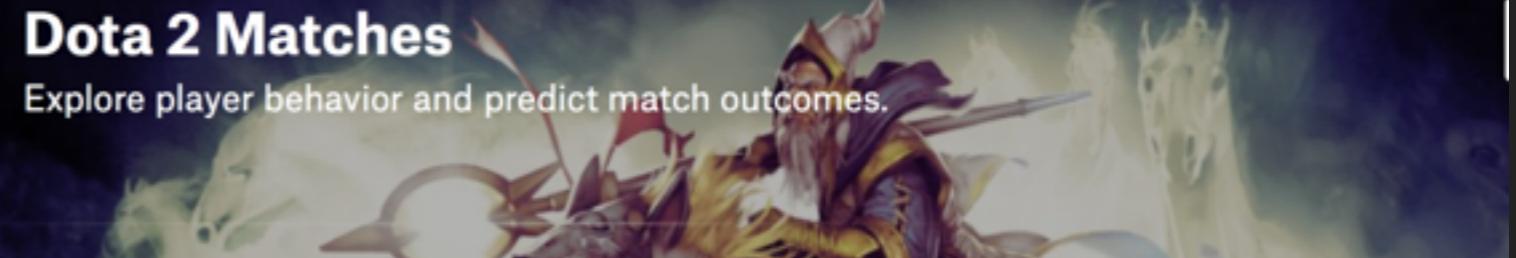
5 in "Diabetic Retinopathy Detection" competition

Research Engineer @ Google DeepMind

WORK WITH THE COOLEST DATASETS



Deloitte.



OPPORTUNITIES



Passenger Screening Algorithm Challenge

Improve the accuracy of the Department of Homeland Security's threat recognition algorithms

· 6 months to

[Featured](#)[go](#)



Zillow Prize: Zillow's Home Value Prediction (Zestimate)

Can you improve the algorithm that changed the world of real estate?

· 7 months to

[Featured](#)[go](#)



Data Science Bowl 2017

Can you improve lung cancer detection?

· 3 months

[Featured](#)[ago](#)



Heritage Health Prize

Identify patients who will be admitted to a hospital within the next year using historical claims data. (Enter by 06:59:59 UTC Oct 4 2012)

· 4 years

[Featured](#)[ago](#)



Flight Quest 2: Flight Optimization, Milestone Phase

Optimize flight routes based on current weather and traffic.

· 4 years

[Ge](#)[ago](#)



GE Flight Quest

Think you can change the future of flight?

· 4 years

[Ge](#)[ago](#)

\$1,500,000
50 teams

\$1,200,000
1,120 teams

\$1,000,000
1,972 teams

\$500,000
1,353 teams

\$250,000
129 teams

\$250,000
173 teams



★ Senior Data Scientist

CarTrawler · Dublin, Ireland - London, UK
posted 7 days ago



★ Data Scientist

OLX Group · Gurgaon
posted 7 days ago



★ Machine Learning Engineer

Cisco · San Jose
posted 9 days ago



★ Experten (m/w) Data Warehos

Unitymedia GmbH · Köln
posted 9 days ago



★ KBC Data Scientist

KBC Bank & Insurance · Leuven or Gent (with)
posted 12 days ago



★ Machine Learning/Deep Learni

SAP · Singapore
posted 19 days ago



SKY
IS THE
LIMIT

TOOLS?



dmlc
XGBoost



pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



dmlc
mxnet



dplyr



R Studio[®]

WHAT WE WILL BE USING

- ▶ Anaconda - Python distribution for Data Scientists
 - ▶ scikit-learn
 - ▶ scipy
 - ▶ numpy
 - ▶ jupyter notebook
- ▶ Algorithm packages
 - ▶ XGBoost
 - ▶ LightGBM
 - ▶ Tensorflow + Keras
 - ▶ Bayesian Optimization
- ▶ Plotting tools
 - ▶ seaborn
 - ▶ bokeh

HOW?



FEATURE ENGINEERING



PARAMETER TUNING



MODEL ENSEMBLE

AND THE TRUE STORY IS



Hard working?

Stay energetic!



Insightful?



Think different!

Blessed?



Damn shitty good luck

QUESTIONS?

