

★ Get unlimited access to the best of Medium for less than \$1/week. [Become a member](#)



As melhores plataformas de Competição para Cientistas de Dados

Conheça plataformas como Kaggle, CodaLab e DrivenData para participar de competições de Data Science



Marlesson Santana · [Follow](#)

Published in Data Hackers · 8 min read · Jul 28, 2019



149



1



Foto por [Andrey Burmakin](#) from Pexels

Este texto é uma **tradução** livre com algumas atualizações da postagem original feita pela Parul Pandey.

Top Competitive Data Science Platforms other than Kaggle

The best way of learning anything is by doing.

towardsdatascience.com

Um agradecimento a autora por permitir a utilização do seu texto.

Como ter experiência na área de Ciência de Dados?

O que você faz depois de ter completado centenas de MOOCs, consumido milhares de livros e anotações e escutado um milhão de pessoas falando sobre sua experiência em Ciência de Dados? **Você começa a aplicar os conceitos.** A única maneira de aplicar conceitos em Ciência de Dados é colocando as mãos nos dados.

A **área de Ciência de Dados** tem uma particularidade interessante, é uma área em que **é possível ganhar experiência sem ter um emprego formal.** Isso é ótimo para quem está começando, pois, nem sempre é fácil ter 20 anos de experiência aos 30 anos para se aplicar em vagas de emprego, e também é fundamental para os mais experientes se manterem atualizados com problemas fora do seu dia a dia.

Uma forma de ganhar essa experiência é através das Plataformas de Competição na Área de Dados. Nelas são divulgados problemas em

diferentes áreas, muitos são problemas reais, com dados disponíveis para exploração e um objetivo em foco, sem contar a possibilidade de ganhar premiação e reconhecimento na área. Competições de DS e ML são muito além do que testar algoritmos prontos, envolve todo o processo de entendimento do problema, modelagem da solução, testes de hipóteses e otimização em busca de um objetivo, sem contar que a possibilidade de comparar soluções com diferentes profissionais traz uma experiência única.

Vantagens de participar em competições de ciência de dados

Você tem muito a ganhar e praticamente nada a perder participando dessas competições. Tem benefícios tangíveis e intangíveis como:

- Ótima oportunidade para aprender.
- Ser exposto a abordagens e conjuntos de dados de última geração.
- Networking com pessoas afins. Trabalhar em equipe é ótimo, já que ajuda a pensar sobre um problema de diferentes perspectivas.
- É uma forma de apresentar seu talento para o mundo e ter uma chance de ser recrutado.
- Também é divertido participar e ver como você se sai na tabela de classificação.
- O prêmio é um bônus adicional, mas não deve ser o único critério.



On September 18, 2009, BellKor Pragmatic Chaos officially won the NetFlix competition by a tiebreaker.

O **Kaggle** é uma plataforma bem conhecida para competições de Data Science. É uma comunidade com diversos usuários, desde novatos a especialistas, que publicam as suas soluções e análises em diferentes problemas e fontes de dados disponibilizados na plataforma.

Caso você não conheça o **Kaggle** e nem o básico de competições de ML. Se liga no canal do Mário Filho e no artigo do Paulo sobre o assunto. Nossos Community Managers dão dicas importantes para quem quer começar a competir.

Aula #1 - Resolva o Titanic Como um Campeão do Kaggle



Como participar de uma competição de Machine Learning no Kaggle

Separe 10 minutos do seu tempo e vamos aprender a participar de nossa primeira competição do Kaggle usando o famoso...

paulovasconcellos.com.br

No entanto, além da Kaggle, **existem outras Plataformas de Competição na Área de Dados que valem a pena conhecer e explorar**. Aqui está uma breve visão geral de alguns deles e uma listagem mais completa ao final levantada pela nossa comunidade.

Driven Data

DRIVENDATA

COMPETITIONS

ABOUT →

DRIVENDATA/

BLOG

LOG IN

SIGN UP

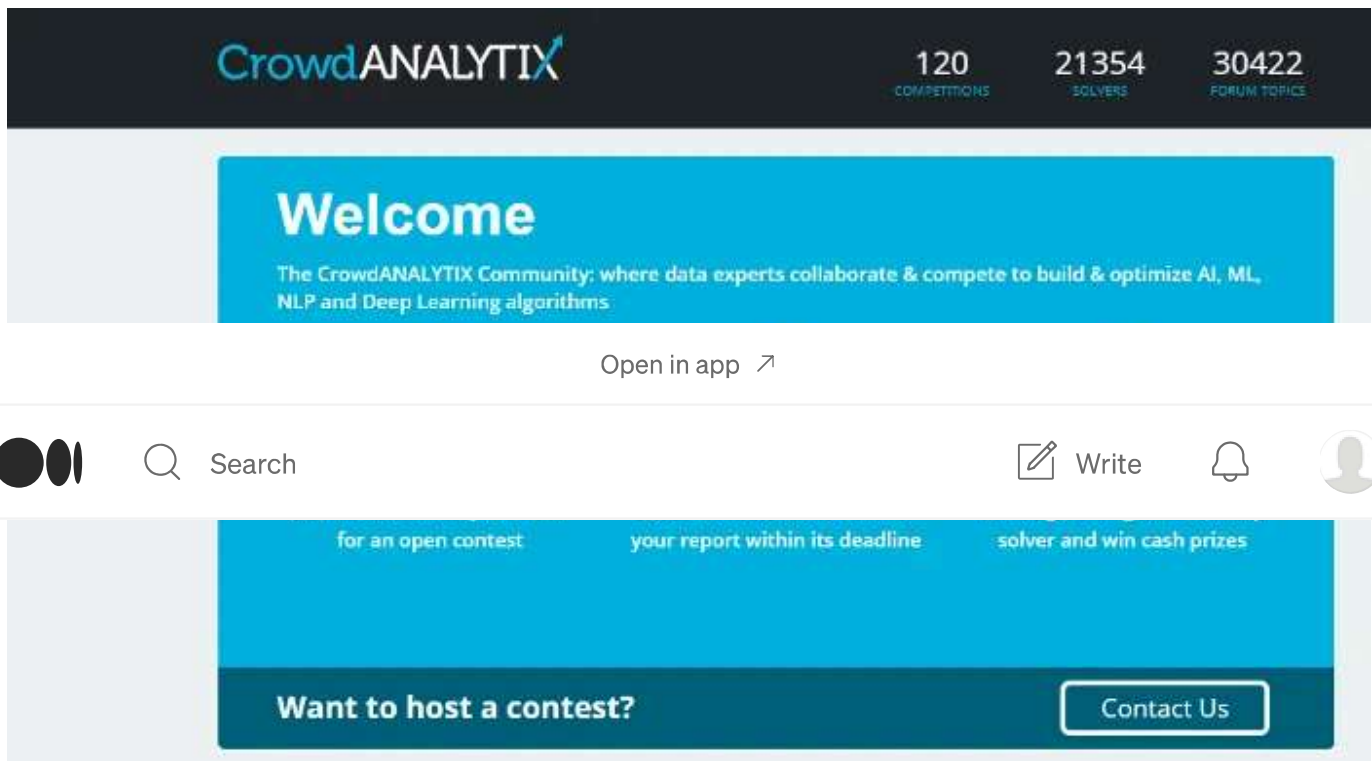
Data science competitions
to build a better world

[I want to join a competition →](#)[I want to run a competition →](#)

A **DrivenData** hospeda competições de ciência de dados para construir um mundo melhor. Eles trabalham com projetos que tenham intersecção entre **ciência de dados e impacto social** em diferentes áreas, como desenvolvimento internacional, saúde, educação, pesquisa, conservação ambiental e serviços públicos.

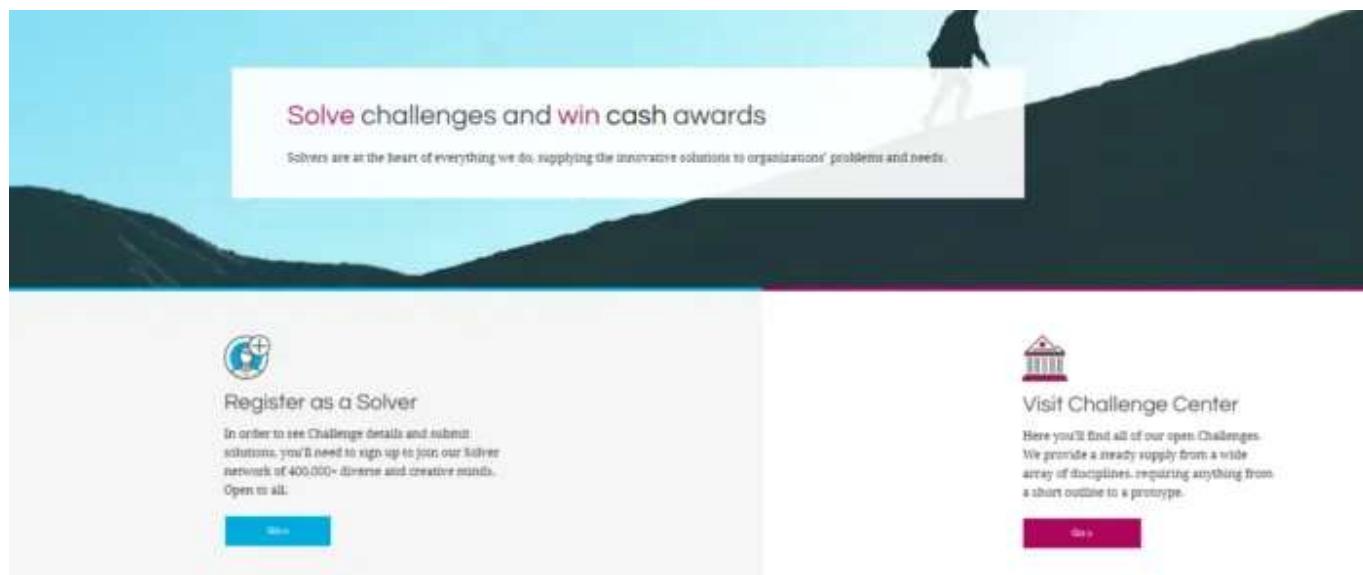
Os conjuntos de dados listados no Driven Data estão relacionados a organizações sem fins lucrativos que vão desde a preservação da vida selvagem até a saúde pública. Assim, **se você quiser aplicar suas habilidades para problemas do mundo real e ainda melhorar o planeta, esta é a plataforma para você.**

CrowdANALYTIX



CrowdANALYTIX é uma plataforma *crowdsourcing* que converte desafios e problemas de negócios em competições. A CrowdANALYTIX Community colabora e concorre para construir e otimizar algoritmos de IA, ML, NLP e Deep Learning. A plataforma também hospeda um blog da comunidade que tem ótimos recursos, incluindo entrevistas e materiais de referência.

Innocentive



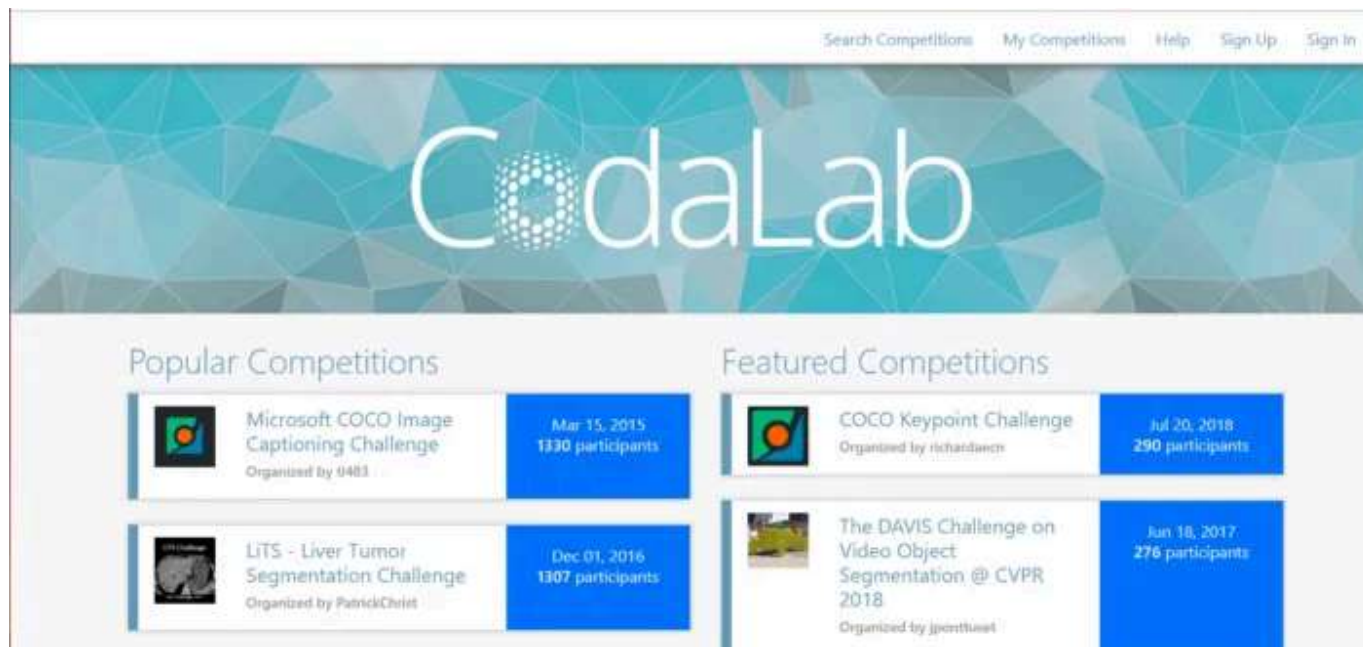
A **InnoCentive** se concentra principalmente em problemas relacionados às ciências da vida, mas também tem outras competições interessantes. Aqui, os *Solvers* contribuem para enfrentar alguns dos problemas mais prementes do mundo, desde facilitar o acesso à água potável a nível doméstico até dispositivos solares passivos projetados para atrair e matar mosquitos portadores de malária. Desafios são problemas reais que exigem concentração sustentada, pensamento crítico, pesquisa, criatividade e síntese de conhecimento. Desenvolver uma solução é incrivelmente gratificante e um treino mental inigualável.

TunedIT



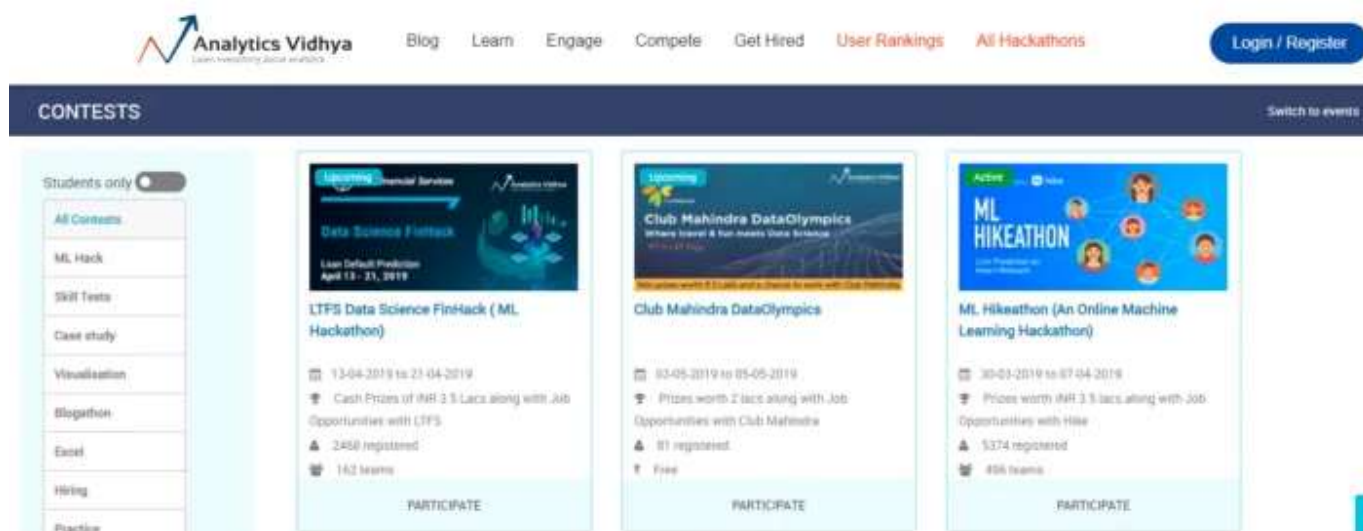
TunedIT começou como um projeto de doutorado científico realizado na Universidade de Varsóvia. O objetivo era ajudar os cientistas de mineração de dados a conduzir experimentos repetíveis e avaliar facilmente os algoritmos orientados por dados. A parte de pesquisa foi complementada mais tarde com a plataforma **TunedIT Challenges** para hospedar competições de dados — para fins educacionais, científicos e de negócios.

Codalab



O **Codalab** é uma plataforma de código aberto que permite que pesquisadores, desenvolvedores e cientistas de dados colaborem, com o objetivo de avançar em campos de pesquisa nos quais o aprendizado de máquina são usados. O CodaLab ajuda a resolver muitos problemas comuns na área de pesquisa orientada a dados por meio da sua comunidade online, onde as pessoas podem compartilhar planilhas e participar de competições. É possível participar de uma competição existente ou sediar uma nova dentro da plataforma.

Analytics Vidhya



O **Analytics Vidhya** fornece um portal com bastante conteúdo para profissionais de *Analytics* e Data Science. Além de fornecer ótimos recursos para o aprendizado na área de Dados, também hospeda os *hackathons*, que são problemas reais da indústria sendo lançados na forma de concursos. Você pode participar dos desafios ou patrocinar um *hackathon*. A maioria das empresas que organizam o *hackathons* no Analytics Vidhya também oferecem oportunidades de emprego para os melhores marcadores.


CrowdAI

crowdAI

ChallengesKnowledge BaseJob BoardSign upLog in


Challenges

All 12Active 14Completed 38




LifeCLEF 2019 Plant
Image-based plant identification on Amazonian flora
23 days left

24901280
ViewsParticipantsSubmissions



ImageCLEF 2019 VQA-Med
23 days left

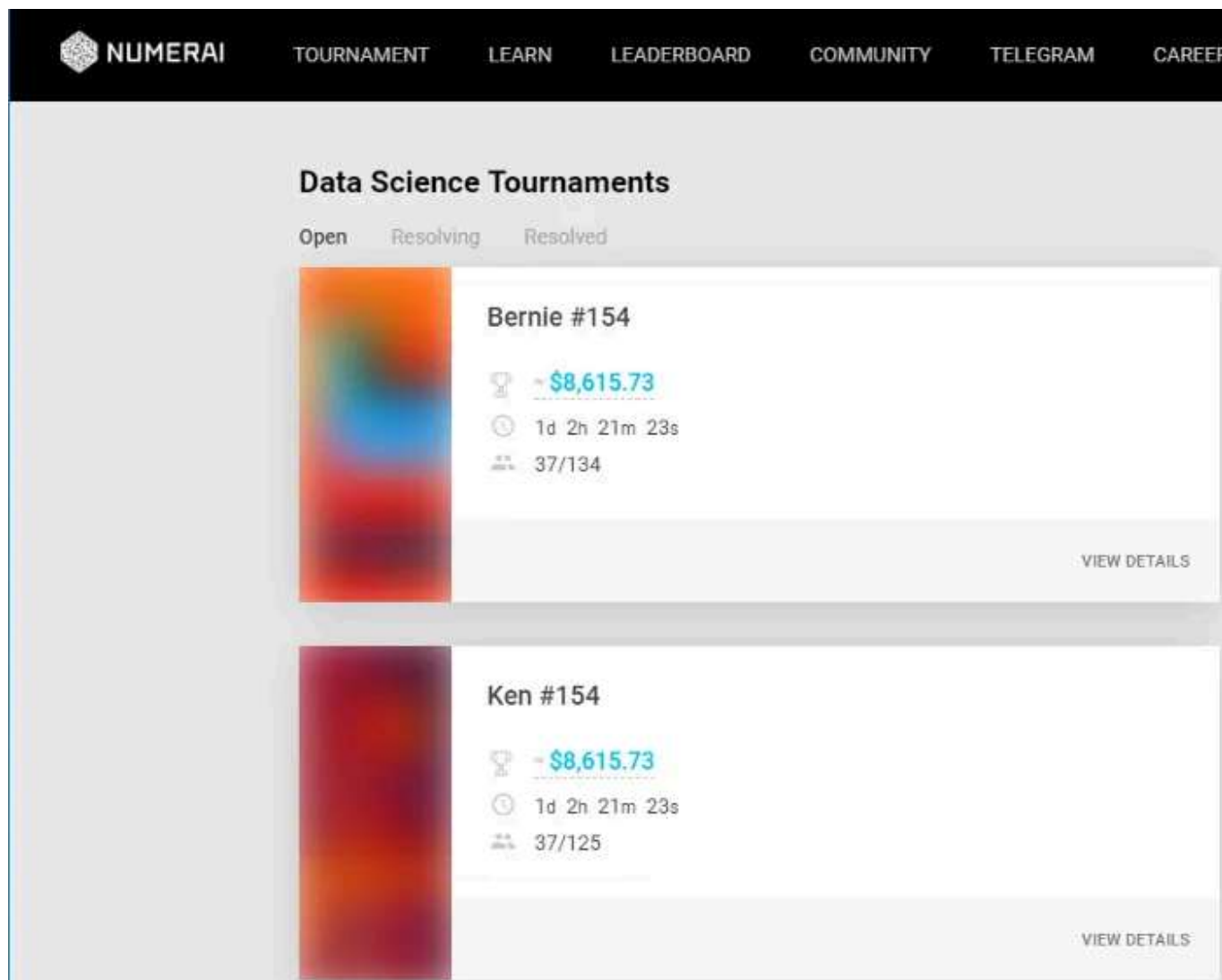
2432940
ViewsParticipantsSubmissions



ImageCLEF 2019 Tuberculosis - Severity scoring

A plataforma de desafio de ciência de dados crowdAI hospeda vários desafios a cada ano. Os desafios abrangem problemas de classificação de imagens, reconhecimento de texto, aprendizado de reforço, ataques adversários, segmentação de imagens, otimização de alocação de recursos e muitas outras áreas em vários domínios. Eles receberam mais de US \$ 100 mil da Amazon e da Nvidia por seu desafio de 2017 chamado “Learning to Run”.

Numerai



Numerai é um fundo de hedge administrado por uma rede de cientistas de dados. Ele realiza uma **competição de ciência de dados a cada semana** que alimenta um fundo de hedge real, então são fornecidos dados criptografados toda semana para seus participantes, que então enviam suas previsões afim de ganhar a criptomoeda da plataforma. Numerai então cria um meta-modelo de todas as suas submissões e faz investimentos.

Os cientistas de dados apresentam suas previsões em troca do potencial de ganhar algum **Numeraire**, que é a criptomoeda utilizada pela plataforma.

Tianchi

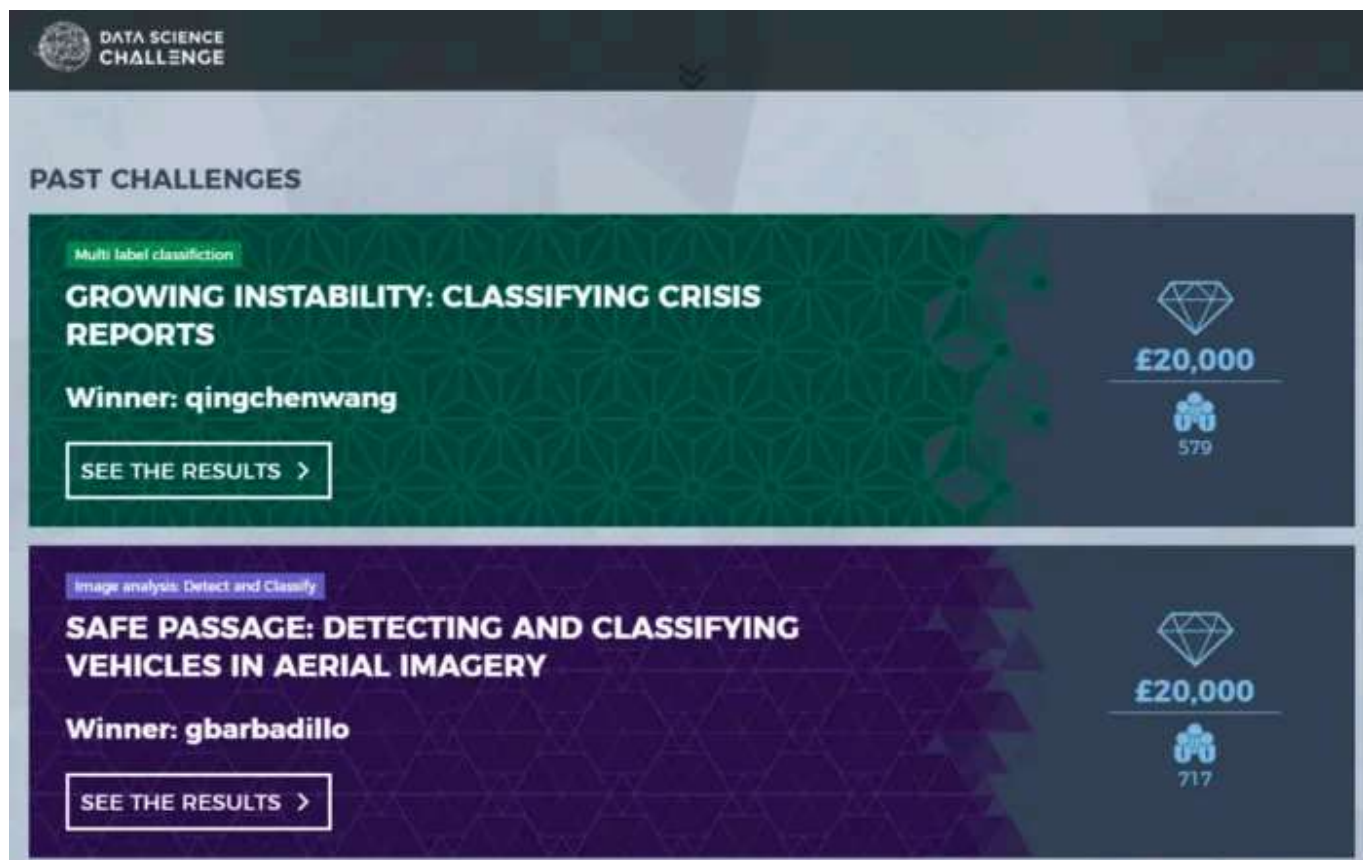
Tianchi Big Data Competition
Big data and distributed computing resources, Cutting-edge solutions for real-world applications.

Active | Algorithm | Innovative | Program | Getting Started | Visualization | CREATE @

Competition Name	Overview	Price	Teams	Season1	Status
CIKM 2019 EComm AI: Efficient User Interests Retrieval	Overview: Efficient and Novel Item Retrieval for Large-scale Online Shopping Recommendation Sponsors: CIKM2019, DAMO, Alibaba, TIANCHI	\$25000	751	2019-08-15	In Progress
CIKM 2019 EComm AI: User Behavior Diversities Prediction	Overview: Predicting User Behavior Diversities in A Dynamic Interactive Environment Sponsors: CIKM2019, DAMO, Alibaba, TIANCHI	\$25000	1111	2019-08-15	In Progress

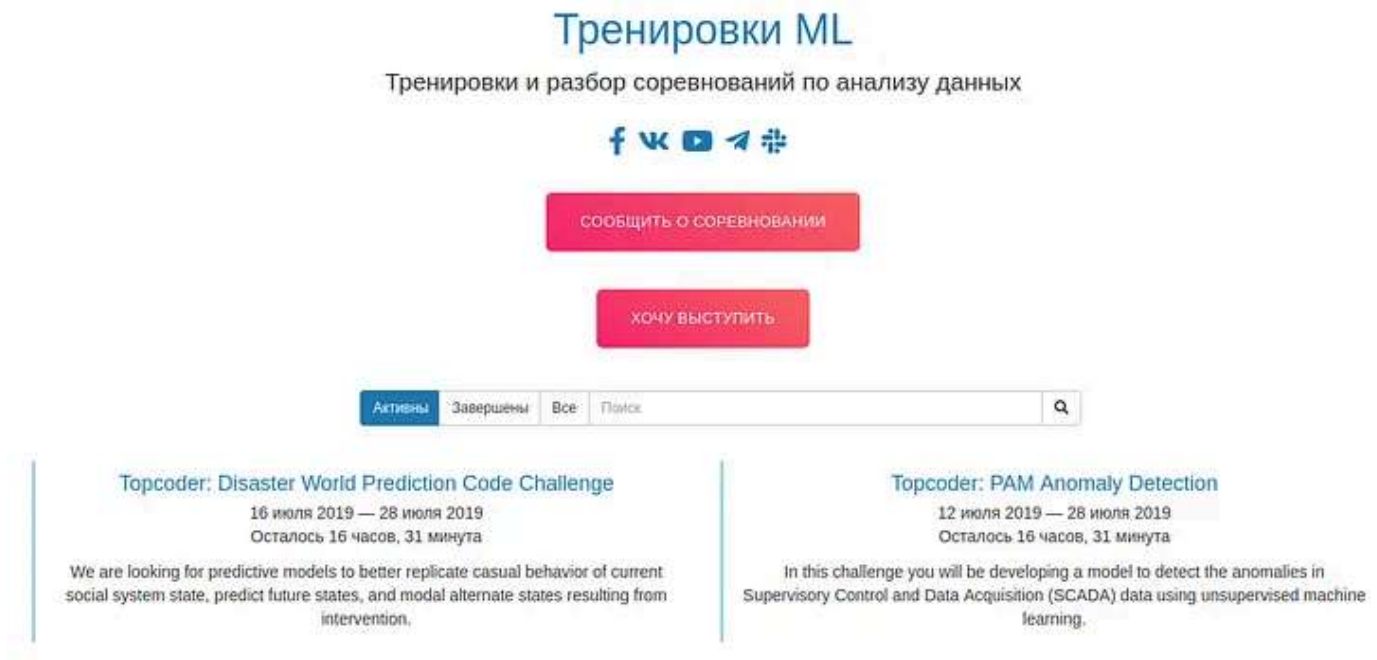
Tianchi é uma plataforma de competição de dados da **Alibaba Cloud** e se assemelha a Kaggle de várias maneiras. É uma comunidade na qual centenas de milhares de cientistas de dados cooperam entre si e se conectam com empresas e governos globalmente para resolver os problemas de negócios mais difíceis em todos os setores.

DataScienceChallenge



Os desafios do DataScienceChallenge são patrocinados pelo Laboratório de Ciência e Tecnologia de Defesa (Dstl), assim como vários outros departamentos do governo do Reino Unido, incluindo o Escritório do Governo para Ciência, SIS e MI5. Os desafios são projetados para encorajar as mentes mais brilhantes da ciência de dados a ajudar a resolver problemas do mundo real.

Тренировки ML (ML Trainings)



Por último o **Тренировки**, site russo que *indexa* várias páginas de competição e centraliza essa informação no seu site de buscas. Ele lista competições tanto de plataformas de competição, quanto de sites específicos que alguma empresa cria para os desafios. O buscador é simples e não precisa saber Russo para usar e pesquisar a competição que mais te agrada.

Otras Plataformas

Além das Plataformas apresentada na postagem, os Data Hackers Gilberto e Mário Filho, Grandmasters no Kaggle, levantaram uma lista enorme de plataformas que disponibilizam competições na área de dados. Confere ai:

- Kaggle — <https://www.kaggle.com/competitions>
- Drive Data — <https://www.drivendata.org/>
- Тренировки ML — <http://mltrainings.ru/>

- **Analytics Vidhya** — <https://datahack.analyticsvidhya.com/>
- **NumerAI** — <http://www.numerai.com/>
- **CrowdANALYTIX** — <https://www.crowdanalytix.com/community>
- **TopCoder** — <https://www.topcoder.com/challenges>
- **TunedIT** — <http://tunedit.org/challenges>
- **CodaLab** — <https://competitions.codalab.org/>
- **Tianchi** — <https://tianchi.aliyun.com/competition/gameList/activeList>
- **DataScience Challenge** — <https://www.datasciencechallenge.org/>
- **HackerRank** — <https://www.hackerrank.com/categories/ai/machine-learning>
- **Grand Challenge** — <https://grand-challenge.org/challenges/>
- **hahub** — <https://chahub.org/>
- **crowdAI** — <https://www.crowdai.org/challenges>
- **KDD** — <https://www.kdd.org/kdd-cup/>
- **Signature** — <https://signate.jp/competitions>
- **DC Jingsai** — <https://www.dcjingsai.com/common/cmptIndex.html>
- **IA Challenger** — <https://challenger.ai/>
- **Kesci** — <https://www.kesci.com/home/competition>
- **Halite** — <https://halite.io/>
- **Codi Game** — <https://www.codingame.com/multiplayer/bot-programming/coders-strike-back>
- **INNOCentive** — <https://www.innocentive.com/our-solvers/>

Caso você conheça alguma outra plataforma que não esteja na listagem acima, manda um *pull-request* no repositório https://github.com/marlesson/list_of_ml_competitions que atualizo a postagem.

Calendar of machine learning competitions

This is a way to collectively solve a real business task using gamification techniques and find the best solution in a...

ods.ai

Não importa o seu nível profissional, competições na área de dados é uma das melhores formas de estudar e aprender coisas novas. Não tenho medo de competir.

Machine Learning

Data Science

Estudo

Dados Abertos

Competition

Some rights reserved 



Written by Marlesson Santana

[Follow](#)

669 Followers · Editor for Data Hackers

Data Scientist and Machine Learning Researcher

More from Marlesson Santana and Data Hackers



Marlesson Santana

Utilizando o Scrapy do Python para monitoramento em sites de...

Introdução ao Scrapy do Python

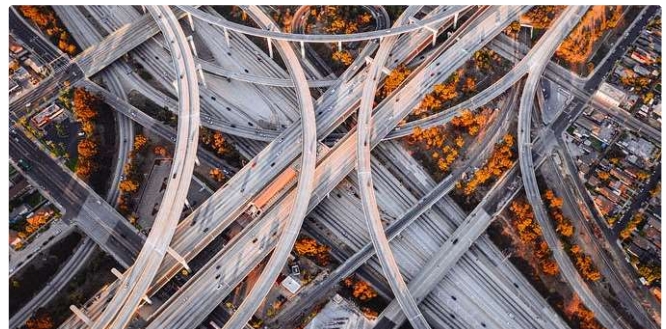
9 min read · Sep 24, 2017



1.1K



10



Caio Estrella in Data Hackers

Pandas: combinando data frames com merge() e concat().

Diferentes estratégias para combinar tabelas

7 min read · May 27, 2020



265



4



```
receita <- c(100,50,11,5,40,140,200)
receita_lizado <- rescale(receita)
frame(formula = rescale(receita),
       formula = (receita - min(receita)) / (max(receita) - min(receita))

formula s_formula
8717949 0.48717949
8076923 0.23076923
8076923 0.03076923
0000000 0.00000000
7948718 0.17948718
9230769 0.69230769
0000000 1.00000000
```

 Arthur Lamblet Vaz in Data Hackers

Normalizar ou padronizar as variáveis?

Um dos processos muito rotineiros para um Data Science é "colocar as variáveis na..."

5 min read · Jun 30, 2019



302



2




469



4



 Marlesson Santana in Data Hackers

Deep Learning para Sistemas de Recomendação (Parte 1)—...

Introdução a Sistemas de Recomendação e Deep-RecSys

11 min read · Dec 6, 2018

See all from Marlesson Santana

See all from Data Hackers

Recommended from Medium





Kallol Mazumdar in ILLUMINATION

I Went on the Dark Web and Instantly Regretted It

Accessing the forbidden parts of the World Wide Web, only to realize the depravity of...

8 min read · Mar 13, 2024



17.4K



317



Karolina Kozmana

Common side effects of not drinking

By rejecting alcohol, you reject something very human, an extra limb that we have...

10 min read · Jan 21, 2024



31K



830



Lists



Predictive Modeling w/ Python

20 stories · 1150 saves



Practical Guides to Machine Learning

10 stories · 1386 saves



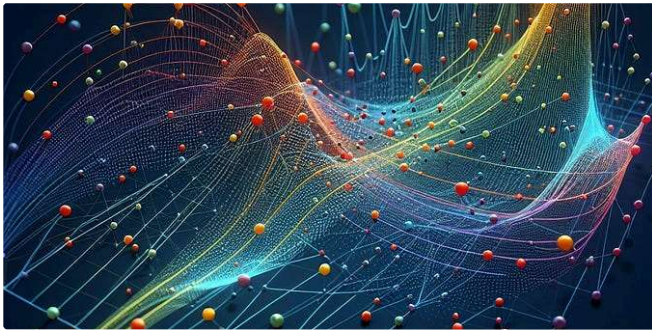
Natural Language Processing

1429 stories · 921 saves



data science and AI

40 stories · 146 saves



Tim Sumner in Towards Data Science

A New Coefficient of Correlation

What if you were told there exists a new way to measure the relationship between two...

10 min read · Mar 31, 2024



3.4K



41



Hazel Paradise

How I Create Passive Income With No Money

many ways to start a passive income today

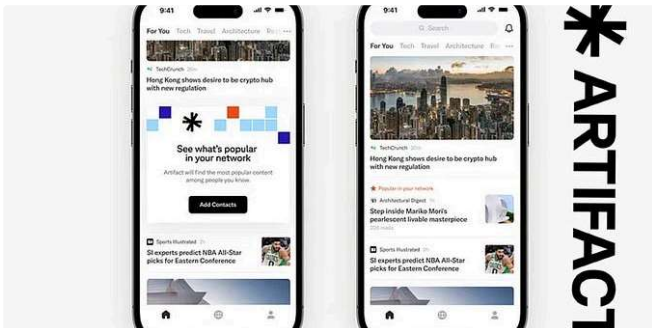
5 min read · Mar 27, 2024



10.1K



238



Gowtham Oleti

Apps I Use And Why You Should Too.

Let's skip past the usual suspects like YouTube, WhatsApp and Instagram. I want t...

11 min read · Nov 14, 2023



22K



409



Unbecoming

10 Seconds That Ended My 20 Year Marriage

It's August in Northern Virginia, hot and humid. I still haven't showered from my...

★ · 4 min read · Feb 16, 2022



78K



1105



See more recommendations