

# **kaggleの始め方**

## **~ 理論から実践へ ~**

# Agenda

- kaggleとは
- なんでやるの？
- どんな感じの流れ？
- やってみた

# kaggleとは？

世界最大の機械学習・データ分析のコンペを主催する  
プラットフォーム

つまり

# **データサイエンティストの 世界最強を決める大会**

# kaggleの規模

- ユーザ数: 50万以上
- 国: 190カ国以上

らしい(´・ω・`)

**なぜこんなことをやるの？**

# 理論から実践へのトレンドの 変化

昔

- 理解する



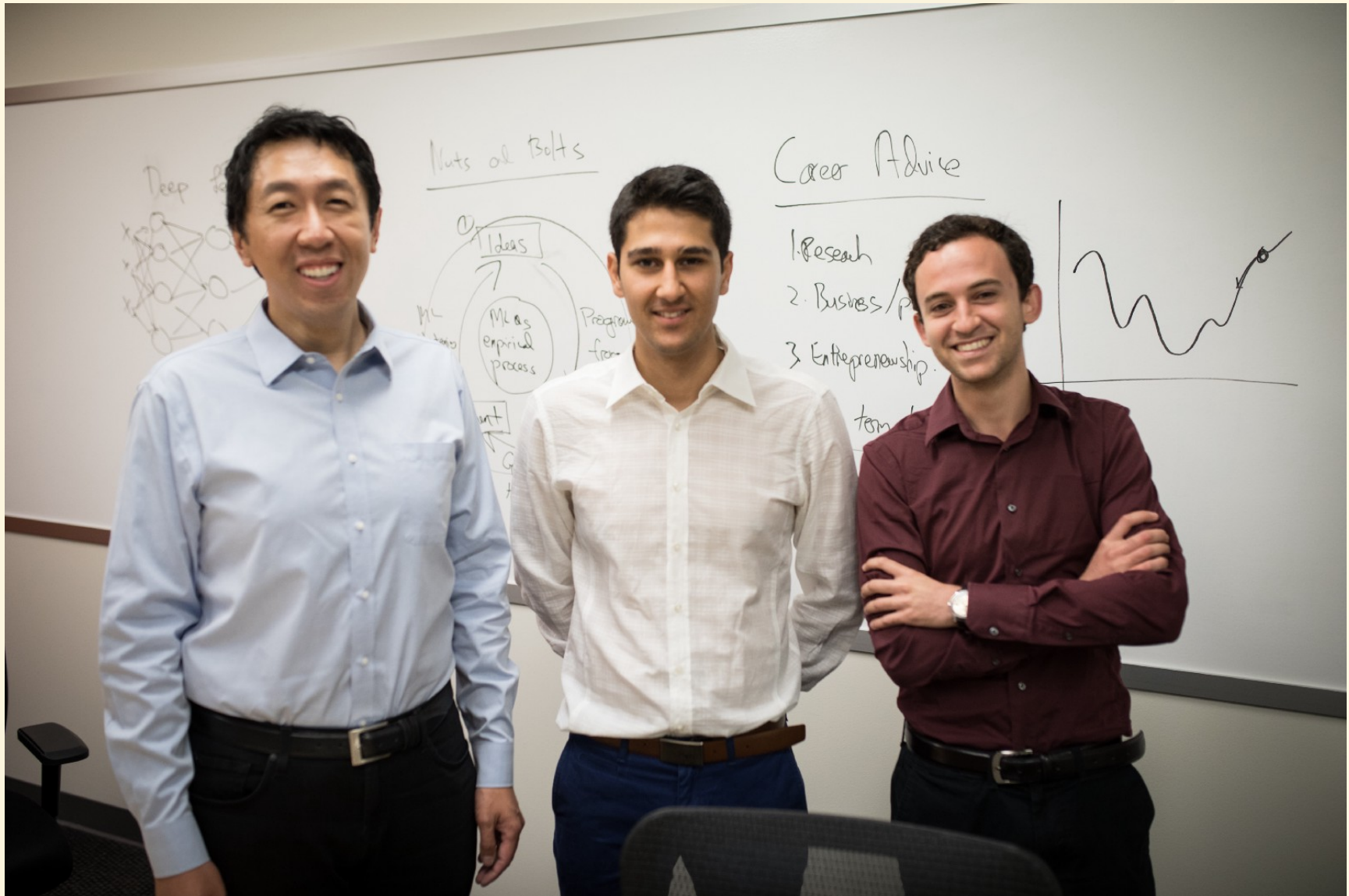
今

- 実践・役立つ



# 実践を重視している具体例

- kaggle
  - コンペ
  - ノウハウの共有（コード・ディスカッション）
  - データの共有・公開
- [fast.ai](https://fast.ai)
  - deep learning for coders（開発者の深層学習）
  - 理論より実践、SOA（state of art: 最先端）



From: [deeplearning.ai: Announcing new Deep Learning courses on Coursera](https://deeplearning.ai/announcing-new-deep-learning-courses-on-coursera/)

# AI社会による生活の向上

by Andrew Ng (AIや機械学習の有名な教授・教師)

*“ I hope we can build an AI-powered society that gives everyone affordable healthcare, provides every child a personalized education, makes inexpensive self-driving cars available to all, and provides meaningful work for every man and woman. An AI-powered society that improves every person's life.*

”

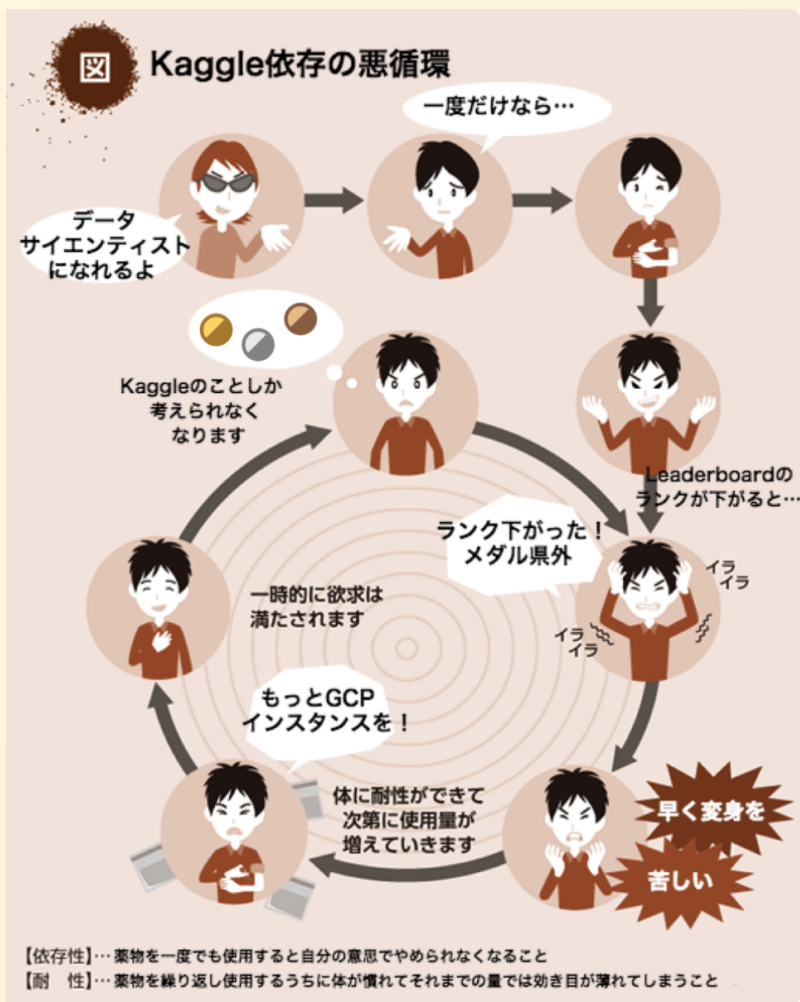
# kaggleの大まかな流れ

1. ホスト（企業など）がコンペを主催する
  1. データを準備
  2. 問題を定義する
2. 参加者は様々な手法を使ってベストなモデルを構築し、予測を提出する <= スコアやランキングが分かる
3. コンペの期限の後、ホストは精度が高い予測を提出した入賞者に賞金を払う

# kaggleをなんでやるのか？

- 個人として
  - 様々なデータに触れられる（企業が実データを提供してくれる。レアい）
  - 他の参加者から学べる
  - 入賞すれば賞金 + 良い仕事をGET！
  - 楽しい
- 主催者側として
  - 世界中のデータサイエンティストが問題解決の手法を試行錯誤してくれる

# kaggle依存の副作用



コンペが始まると仕事しなくなる人たち

# kaggleを実際にやってみた

## 1. コンペを選ぶ

[image:95BD2063-FE62-4019-97B7-809AC6FAAB99-77261-0003D3AF326AB4A3/スクリーンショット 2018-09-21 13.51.37.png]

最近始まったばかりのコンペ(´・ω・`)

[image:B820416D-F306-45AB-A39A-641C9523DEC5-77261-0003D3B25B7989CD/スクリーンショット 2018-09-21 13.52.08.png]



## 2. コンペの内容を読む

1. 概要: 大まかにわかると色々アイディアが出て楽しい
2. 評価指標: これが一番大事ってわかんかね(´・ω・`)
3. 賞金: できればほしいよね
4. 期限: 時間厳守
5. データ: だいたいCSVファイル

### 3. 他の参加者から学ぶ

1. コード (kernel)
2. ディスカッション (discussion)

[image:2093AB9F-EB2D-4D8B-B146-B3103F9F7F13-77261-0003D3D02D70BB66/スクリーンショット 2018-09-21 13.54.32.png]

いろんな人がコードを載せてくれているので助かる

[image:EBB850C6-4AE6-45F6-BAE7-EEF4EC631099-77261-0003D3E7C0E6ACDE/スクリーンショット 2018-09-21 13.56.19.png]

see: [Simple Exploration+Baseline - GA Customer Revenue | Kaggle](#)

## 4. 他の参加者の方法を真似てみる

コードをパクってローカルPCで実行するだけの簡単なお仕事( `・ω・` )

[image:C9EAFF4D-17BC-4209-8394-E40C5EE97F2A-77261-0003D40FAA0F03D0/スクリーンショット 2018-09-21 13.58.52.png]

## 5. 助け合う <= New!

ちょうどライブラリのバージョンで上手く動作しなかったなので、上手くいった方法を教え合う(´・ω・｀)

[image:CB1189CA-FFBD-41B5-81CE-0EF8D4B9A640-77261-0003D41DE2AFBA59/スクリーンショット 2018-09-21 13.59.56.png]

## 6. めんどくさいので人のコードを forkする

[image:406F7B1A-493D-4ECE-92B3-  
4989F79F191A-77261-000424BDD459F0E4/スク  
リーンショット 2018-09-23 21.21.15.png]

[image:13E7887B-740C-41EA-A5B9-  
56748BCC0E15-77261-000424B067F7AE18/スク  
リーンショット 2018-09-23 21.19.58.png]

[image:4041FB8A-3D27-43C4-9EFA-  
A9BC622B2D1A-77261-000424C448186A69/スク  
リーンショット 2018-09-23 20.45.07.png]





## 7. 予測を提出する

[image:FAC748F4-6F51-405E-B654-2B82A01C3B07-77261-000424CC4B606927/スクリーンショット 2018-09-23 21.09.21.png]

## 8. スコアとランクを確認

689位 (全1,031チーム)

ちーん(´・ω・｀)

689	new	data ninja	 	1.7342	1	1m
-----	-----	------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------	---	----

**Your Best Entry ↑**

Your submission scored 1.7342, which is not an improvement of your best score. Keep trying!

# まとめ

TODO

# 參考資料

- [Kaggle - Wikipedia](#)
- [What is Kaggle, Why I Participate, What is the Impact?](#)
- [fast.ai · Making neural nets uncool again](#)
- [deeplearning.ai: Announcing new Deep Learning courses on Coursera](#)