機械学習
データ分析コンペについて

岩政公平
iwamasa@morphometrics.jp
九州大学理学部生物学科
数理生物学研究室

データ分析コンペティションとは

**ある課題に対して統計学や機械学習などの
手法を用いて予測精度を競い合う**

どんな予測を行うか

SIIM-FISABIO-RSNA COVID-19 Detection - Kaggle
胸部X線写真からCOVID-19の疾患箇所を予測する(画像, 物体検出)
(https://www.kaggle.com/c/siim-covid19-detection)

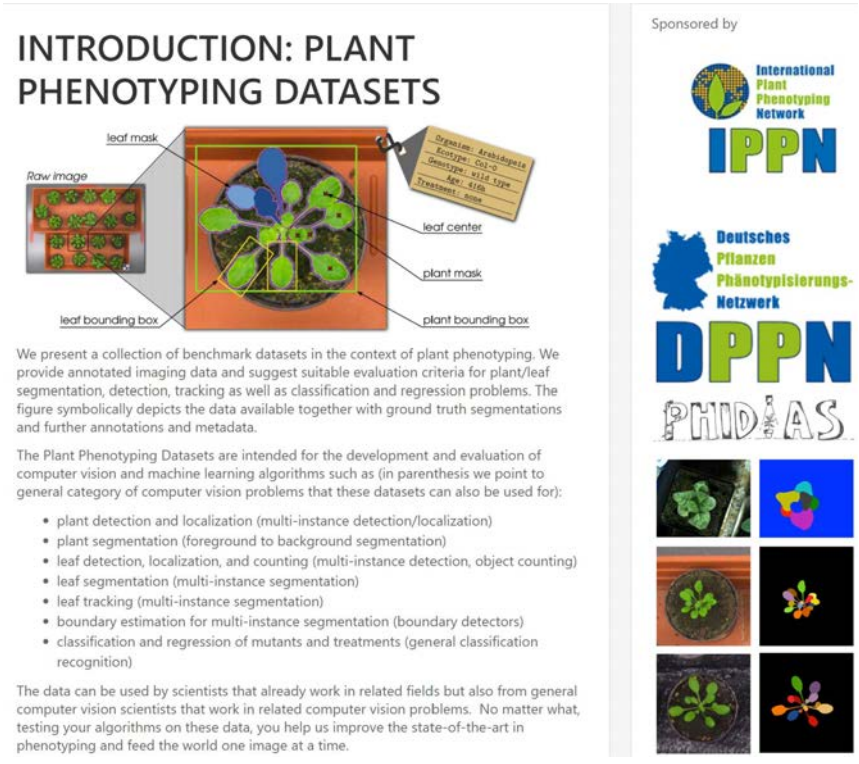
Riiid Answer Correctness Prediction - Kaggle
あるユーザーが問題を正解できるか行動履歴から予測する(時系列, 分類)
(https://www.kaggle.com/c/riiid-test-answer-prediction)

Cornell Birdcall Identification - Kaggle
鳥の鳴き声からその鳥の種を予測する(音声, 分類)
(https://www.kaggle.com/c/birdsong-recognition)

マイナビ × SIGNATE Student Cup 2019: 賃貸物件の家賃予測
所在地や間取りから賃貸物件の価格を予測する(テーブル, 回帰)
(https://signate.jp/competitions/182)

など

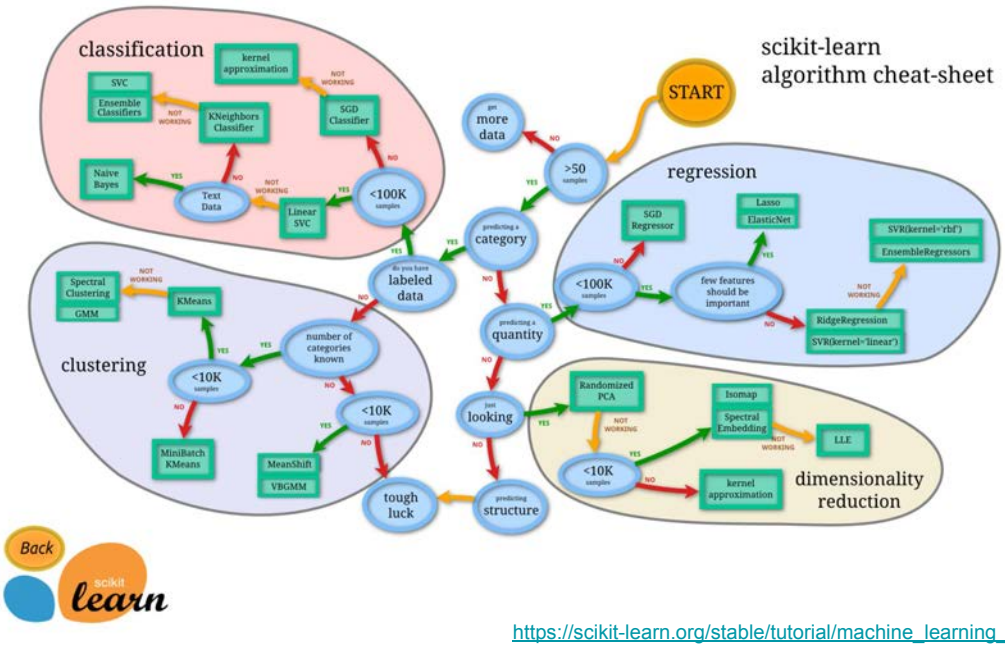
どんな予測を行うか



https://www.plant-phenotyping.org/datasets-home

どんな魅力があるのか
いろんな手法(モデル)を知れる

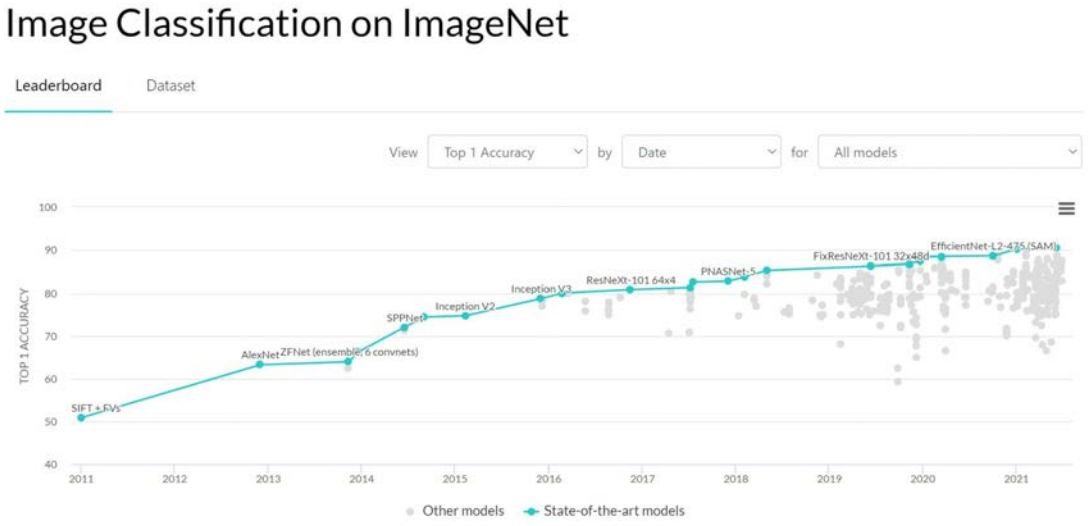
- 機械学習のモデルは様々ある



https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

どんな魅力があるのか
いろんな手法(モデル)を知れる

- より高精度なモデルが毎年提案される



https://paperswithcode.com/sota/image-classification-on-imagenet

どんな魅力があるのか
いろんな手法(モデル)を知れる

- コンペを通して様々なモデルを実際に試してみることができる
  - 新しく提案されたモデルが必ず精度がいいとは限らない
- 他の参加者とディスカッションすることができて互いにモデルの精度の良し悪しや実装コードを共有することができる
- 専門外の分野について知ることができる

モデル以外にも、

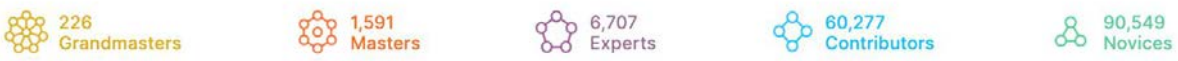
- データ加工の仕方や評価データの作成などの手法を知ることができる
  - 1億行あるテーブルデータや1000万枚ある画像データなど機械学習以外の部分で工夫しないといけないコンペもある
  - データの可視化の方法なども学べる(可視化大事)
  - プログラミングスキルや英語力も上達するかも

どんな魅力があるのか
賞金・メダルがもらえる

- 上位に入賞すると賞金やメダル(称号)がもらえる
- Kaggleの場合は上位のチームにメダルが与えられる

Goldメダル	Silverメダル	Bronzeメダル
上位10チーム+α	上位5%	上位10%

- このメダルを集めると称号が与えられる(2021/07/19現在)



どんな魅力があるのか
楽しい

- 競技性が高く、順位が伸びることが楽しい
  - ランキングで上位になるのは単純に嬉しい
  - コンペ期間中は常にランキングが更新されるためハラハラ感が楽しめる
- データを見てこの処理をすると精度が向上しそうと考える作業が楽しい
  - データの勘所が捉えられれば一気に精度が向上することもしばしば
- 知らない分野・モデルに触れることができる
  - 日本で行われるコンペもあるため初めてみたい人はここがいいかも
  - atmaCup: https://www.guruguru.science/
  - SIGNATE: https://signate.jp/
  - Nishika: https://www.nishika.com/