

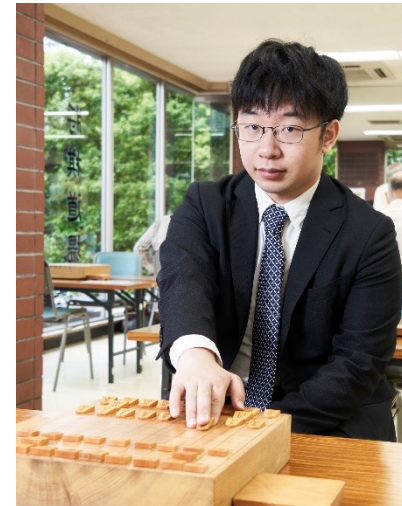
# ファンダメンタルズチャレンジ 3 位解法

谷合 廣紀 (nyoki-mtl)

# 自己紹介

## ■ 谷合 廣紀 (nyoki-mtl )

- 将棋のプロ棋士
- 修士（情報理工学）
- 「楽しいことを全力で」がモットー
  - 将棋楽しい！
  - コンペ楽しい！



# なぜJQuantsコンペに参加したか

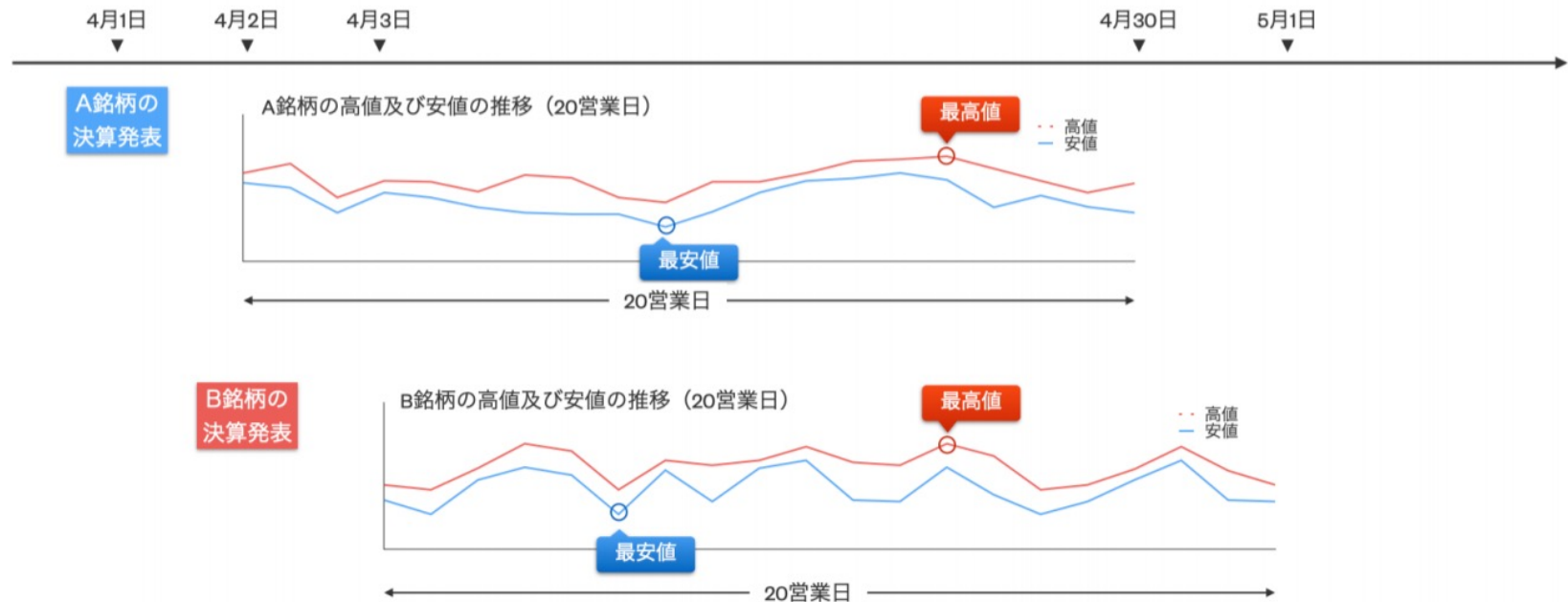
---

- 日本取引所グループが開催
  - APIが用意されていて、実際の日本株データが対象
- チュートリアルが充実
  - EDA、ベースモデルの学習/推論に限らずシステムトレードの情報が網羅的に記述
- システムトレードってなんかカッコいい

楽しそう！

# ファンダメンタルズチャレンジ

## ■ 起点日から20営業日後の最高値と最安値への変動率



## ■ スコアはそれぞれの順位相関係数から計算される

- $score = (\rho_{high} - 1)^2 + (\rho_{low} - 1)^2$

# コンペの特徴

## ■ ランタイムコンペ

- Kaggleのコードコンペのようなもの
- 予測結果ではなく予測コード一式を提出

```
.
├── model                必須 学習済モデルを置くディレクトリ
│   └── ...
├── src                  必須 Python のプログラムを置くディレクトリ
│   ├── predictor.py    必須 最初のプログラムが呼び出すファイル
│   └── ...              その他のファイル (ディレクトリ作成可能)
└── requirements.txt     任意
```

## ■ 実際の日本株の動きで順位が決まる

- 最終評価はコンペ終了後の2021年の3/27~5/15
- PublicLBはコンペ開催以前の2020年の1/1~11/30
  - 正しい暫定順位がわからない！

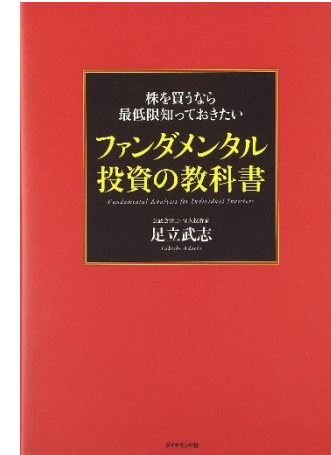
# コンペの方針

---

- 使える時間は一週間程度だったので方針は決め打ち
- トレードの基礎を勉強する
  - 何もわからないと何もわからない
- SIGNATEのランタイム環境はローカルで再現する
  - ランタイム環境周りで詰まるとかなり面倒
- モデルはLightGBM+Optuna
  - アンサンブルやスタッキングをやる余裕はないはず

# ドメイン知識の勉強方法

- チュートリアルに目を通した
  - コンペの目的を理解できる
- 本を読んだ
  - 最低限の用語やトレードの考え方
  - チュートリアルでおすすめされていた一冊
- フォーラムは読み込んだ



# 環境を整える

---

## ■ ランタイム環境は厄介

- メモリや推論時間の制限
- 結果が出るまで時間がかかる
- エラーが出てもデバッグしづらい

## ■ Makefile

- ランタイム環境と同じdocker上に構築
- Train/Validation/Submitすべてコマンド一発
- Submitはランタイム環境で実行できることを確認したのちに提出用ファイル一式を作成する

```
charmer@DESKTOP-2SUSKE3 [3:05:58] ~/Projects/Signate/jpx1  
$ make docker-create-submit-files
```



# 特徴量

---

- 生データはファンダメンタル情報と株価情報
  - ファンダメンタル情報
    - 各銘柄の決算情報
    - 売上高・営業利益・キャッシュフローなど
  - 株価情報
    - 各銘柄の株価情報
    - 始値・終値・高値・安値など
- これらのデータから思いつくだけ特徴量を作成

# ファンダメンタル特徴量

---

## ■ 一般的な財務指標

- PER/PBR/ROE/時価総額/純利益率/自己資本比率…

## ■ それぞれの財務指標の変化率

- 前期と比べて純利益率がどれだけ増加したか/減少したか
- 前期の予測PERと当期の実際のPERとのずれはどのくらいか

## ■ 業種で正規化した財務指標

- 財務指標は業種の中での相対値が重要
- 銘柄の17業種/33業種区分の平均値で割って正規化

# 株価特徴量

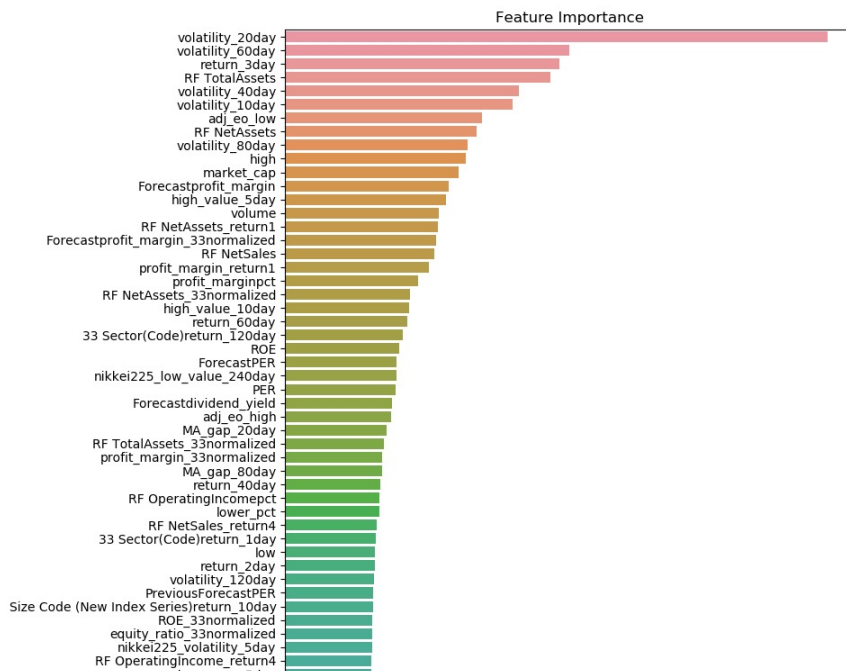
---

- 直近 $n$ (1, 2, 3, 4, 5, 10, 20, 40, 60, 80, 120, 240)営業日における各銘柄/日経平均の株価の
  - リターン
  - ボラティリティ
  - 移動平均線との乖離率など
  - 最高値/最安値
- 決算発表日翌日の制限値幅
  - 予測する最安値の分布の最大値は、制限値幅の最大値に依存
- 業種や日経平均で正規化した株価

# Feature Importance

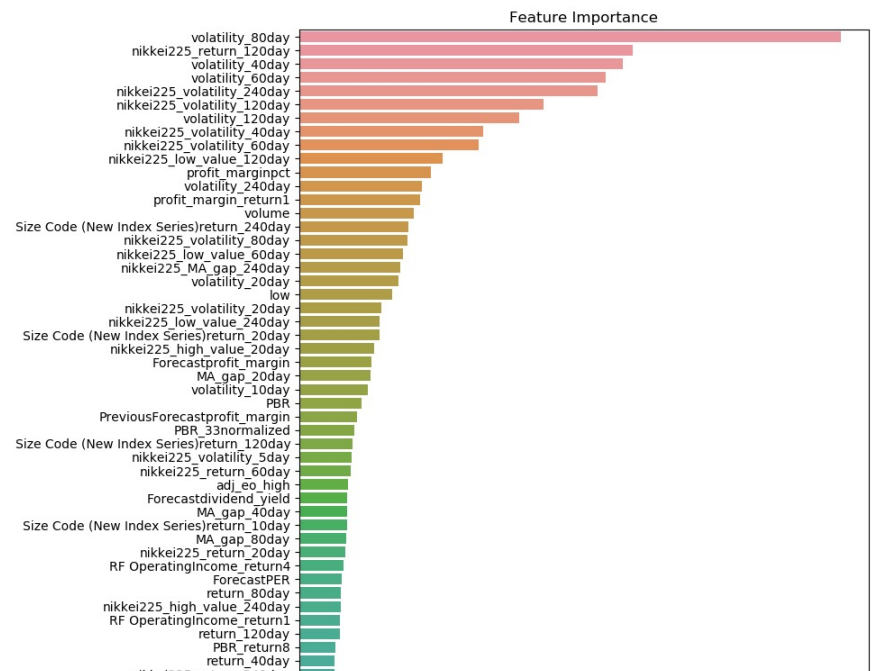
## ■ 高値予測

- ボラティリティ
- 決算発表日の安値/高値
- 総資産、純資産
- 純利益率の来期予想




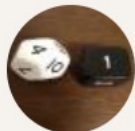
## ■ 安値予測

- ボラティリティ
- 日経平均のボラティリティ
- 日経平均の最安値
- 純利益率の変化率



# 結果

- 暫定順位がわからなかったこともあり  
あまり期待していなかったが最終順位は 3 位に

 3rd	 nyoki-mtl	0.75446
--	---	---------

- ランタイム環境でつまづかないで  
特徴量作成に専念できたのがよかったかも
- 勉強もできて入賞もできるなんて

# 楽しかった！