

＼ Kaggle日記について ／

歴史とか背景とか



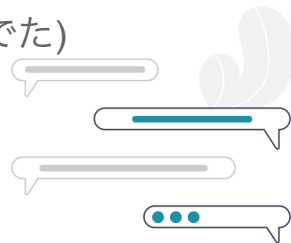
fkubota 

自己紹介



fkubota ([Twitter](#), [Kaggle](#))

- バンドルカードの会社(カンム)で機械学習エンジニア
- Kaggle Expert
- 沖縄出身(東京に出てきて3年半)
- 物理学科出身(強相関電子系、4.2Kの液体ヘリウムに磁性体突っ込んでた)
- プログラミング歴は2年半ぐらい
- 趣味
 - 早起き(4時半起床)
 - Kaggle
 - コーヒー、ビール、ウィスキー
 - 読書、物理、哲学



Kaggle日記って何？

とりあえずどんなものか
簡単に紹介します。

記事があるよ。[\(url\)](#)

初期に コンペ辛かった話

コンペ始めた頃、全然やる
気起きなかったので、
その時抱えていた問題を
ちょっと整理。

だから Kaggle日記始めた

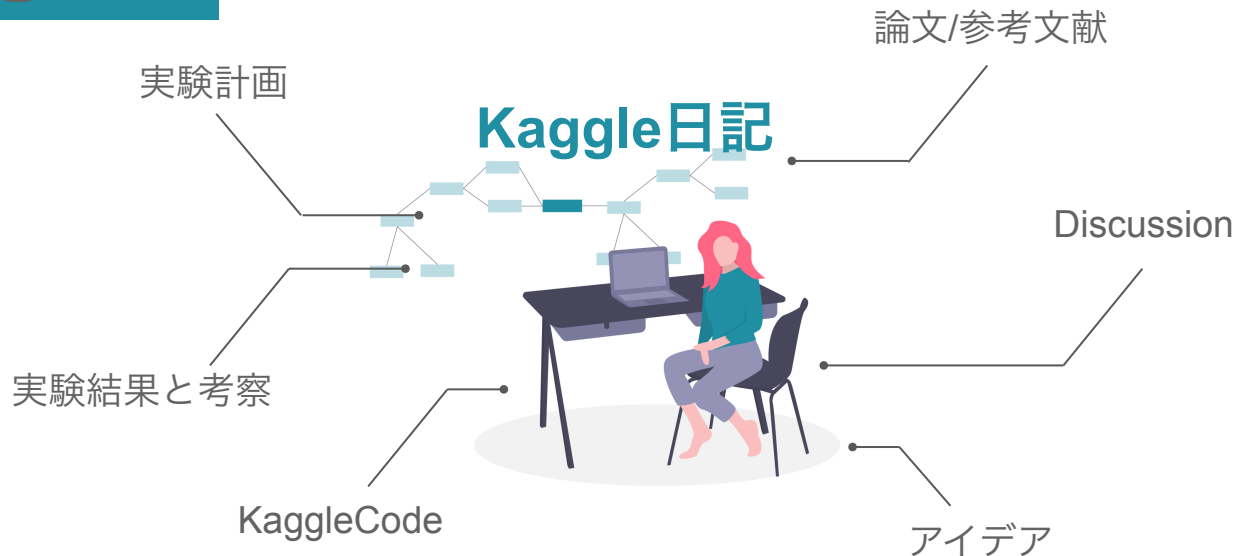
Kaggle日記を始めた理由を
伝えればより使いやすくなるかなと。
そもそもKaggle日記である
必要すらないという話。



Kaggle日記って何？



こんな感じ！



Kaggle参加中に生まれる情報は全て**Kaggle日記**に
例) [鳥コンペのKaggle日記](#)

Kaggleの大変なところ

toooooo long



difficult



何してたか忘れちゃった

命名規則の破綻

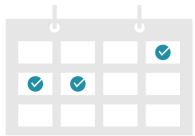
モチベ続かないよ...

アイデアの枯渇

too long問題への対策



Kaggle日記が取り組んでいること



モチベ維持

長いのにほんとうまくいかな
いとか地獄じゃねーか。



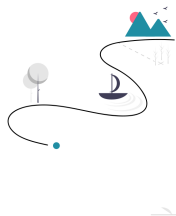
ファイル名で管理しない

どうせ破綻するんでしょ？
なら始めから管理することをあ
きらめようぜ。



脳のメモリで管理しない

3ヶ月もあるんだよ？
覚えられるわけがない。



トレーサビリティ

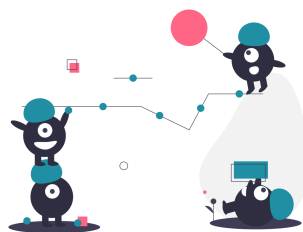
言いたいだけ。

詳細は [「Kaggle日記という戦い方」](#) を見てくださいm(_ _)m

初期にコンペ辛かった話



Kaggleに興味持ち始めた時



興味

人工知能を競いあい
ながら学べるだど！？
やるしかねえ！



完全理解

決定木とSVMを理解した。
え？sklearnですぐ使えるの？
余裕じゃんww



静寂

決定木とSVM試したか
らもうやることないよ？

1438	Ahmad		0.0645	3	2y
1439	fkubota		0.0621	1	2y
Your First Entry ↑ Welcome to the leaderboard!					
1440	Manan Soni		0.0498	1	2y
1441	Saichandra		0.0000	2	2y
1442	Sunil Goda		0.0000	1	2y
1443	yassirbendou		0.0000	1	2y

1サブで終了！
ディスカッションも何書いてる
かわからんしこれ以上は辛い！



まだ...**その時**ではない

その時が訪れるのを勉強しながら待ちました



その時 は一生やってこない



実践はやっぱり違うんだろぅなぁ

お勉強



実践



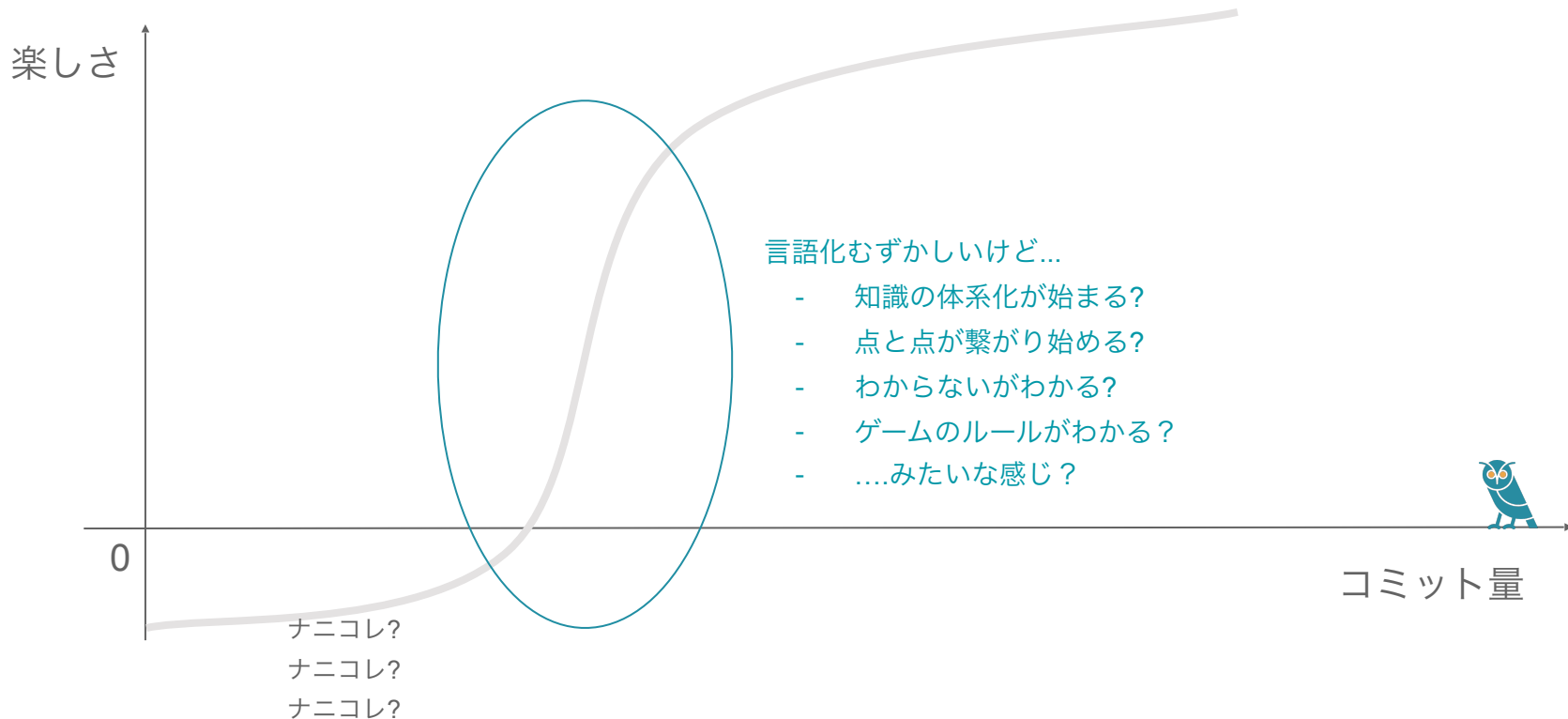
実践でしか学べないことってあるよねえ

- EDA
- リーク
- CVの切り方
- 特徴量選択
- などなど...

わかるけどしんどいもんはしんどい
向いてないのかな？

あくまで僕の経験例なんですが

何か学ぶ時ってだいたいこんな感じ曲線



何が言いたいかと言うと

あくまで僕がそう思ったってだけね！！

楽しさ

ひとまずここをどうにか
乗り越えようと思った

ここで向いていないと判断するの
はもったいないかなあ

0

ナニコレ期

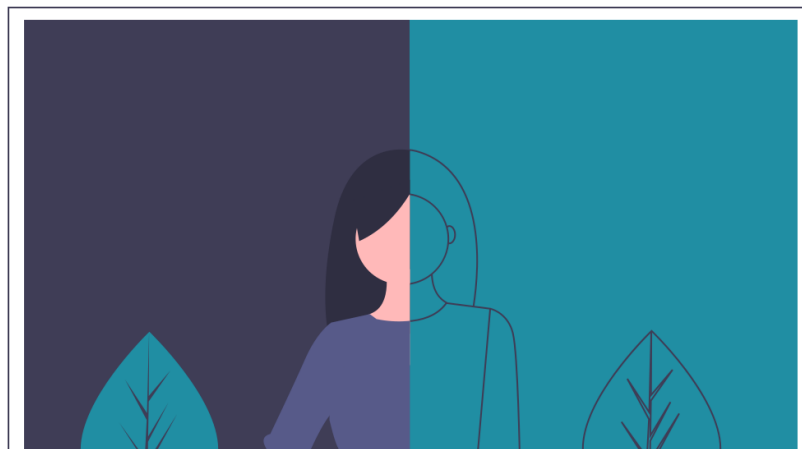


問題が少し明確になって

コンペしんどい
どうしよう問題



ナニコレ期を
どう乗り越えるか問題



だからKaggle日記はじめた



ストレスになりそうなもの

命名規則の破綻

exp_6fold_3.ipynb
exp_5fold_nn_v5.ipynb
exp_5fold_seed42.ipynb
exp_5fold_fix_bug.ipynb
exp_4fold_nn_prepro.ipynb
exp_5fold.ipynb
exp_5fold_nn_postproc_v2.ipynb

作業再開コスト

2日空けちゃって再開しようと思ったけどめんどくさいから明日やろう。

無限ループ

アイデアの枯渇

もうアイデアが尽きました。



再現性がない

A.py動かしてできたa.csv
を使ってB.pyを動かせば再現できるはずだよ。
まあ、できなかったんですけどね:)



outputファイルの管理

この

ultra_super_feature.csv
はどうやって作られたんですかね？

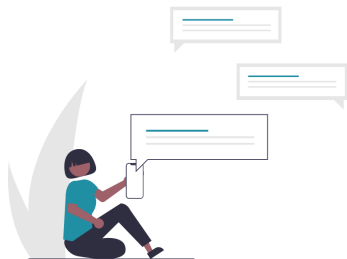


参考文献の管理

参考になりそうな論文やら記事やらを保存したけどどれを読んでどれを読んでないのかわからん。

Todo Doing Done

対策できれば何でも良い



前ページに上げたような問題にできるだけ無理なく対処するために**Kaggle日記**という方法を取りました



ただ、**ちょっと注意**なのですが

- Kaggle日記である必要はないし
- そもそも細かい問題にそれぞれ対処を無理してする必要もないし
- なんならしんどいならKaggleやらない選択肢もある

ので僕の意見は**選択肢の一つぐらいに受け取ってください**

分子コンペの思い出

20

public LB
(score)

銅メダルライン

ここまで貯めていたオリジナルの
アイデアの実装をはじめた。
とにかく実装速度が遅い。

15個程度試して
3個ぐらいが当たった。

最後の7日でアンサンブル
を勉強して試した。

残り2日ぐらい
で初めて
銅メダル圏内に

銅メダル(9%)
フィニッシュ！

この時期はcodeとdiscussionを全部見てとりあ
えず使うというのを行いスコアが上がった。
ただ公開ノートブック未満。

疑問がたまった頃にたまたま
カレーさんの
「Kaggleのチュートリアル」
に出会い完全に理解する

groupby
チョットヨクカラナイ

lightGBM ???

0

pandasって
ナンドス？

oofって何！？

テーブルがなんで
たくさんあるの！？

3ヶ月

時間



初参戦のコンペ！！
Kaggle日記デビュー！！

Thanks :)

