

# 三連単を当てたい

B2 福井駿太

# 背景

- ▶ 三連単を当てたい
- ▶ 三連単は当たらない
  
- ▶ そうだ、pythonしよう！

# 前提知識

- ▶ 三連単（競馬）とは
  - ▶ 1着、2着、3着の馬を、その順に予想する買い方
  - ▶ 最も難易度が高い

# 目標

- ▶ 出馬情報を基に、1着、2着、3着になる確率の高い順に予想する
- ▶ 予想した結果と事前オッズから、最も効率の良い買い方を計算する

# 手段 1

- ▶ 過去に同じコース、距離で行われた競馬の結果をスクレイピングし、csvファイルのデータセットとして格納する
- ▶ 予想したい競馬の出馬表を、同様に格納する
  - ▶ スクレイピング先 : netkeiba.com ( <https://www.netkeiba.com/> )

## 手段 2

- ▶ 過去の結果をもとに回帰分析を行い、結果を予測する
  - ▶ 今回は線形回帰、ランダムフォレスト、SVR（サポートベクター回帰）の三つの手段を用いる

# 結果 1

## ▶ 過去の結果

▶ 17年から21年、中山競馬場芝2500m

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Unnamed	rank	frame	number	name	sex_and_impost	jockey	popularity	odds	stable	weight	course	distance	year	
2	0	1	8	10	イチダイ 牡4	56	内田博	3	4.4	美浦両崎 536(-2)	芝	2500	2017		
3	1	2	8	9	マイネル 牡6	57	津村	7	17.5	美浦高橋 480(-2)	芝	2500	2017		
4	2	3	7	8	サブライ 牡5	57	田辺	4	5.7	美浦池上 3504(+4)	芝	2500	2017		
5	3	4	7	7	ウイング 牡4	56	松岡	5	8.5	美浦手塚 452(-4)	芝	2500	2017		
6	4	4	1	1	エニグマ 牡5	57	福永	1	3.1	美浦戸田 494(+4)	芝	2500	2017		
7	5	4	5	5	デナリ 牡4	56	戸崎圭	2	4	美浦田村 486(0)	芝	2500	2017		
8	6	4	3	3	タケルラ 牡5	57	勝浦	6	14.2	美浦田村 486(+2)	芝	2500	2017		
9	7	4	2	2	デルマサ 牡6	55	シュタ	8	38.2	美浦堀井 476(-10)	芝	2500	2017		
10	8	4	4	4	ヘイロー 牡7	57	吉田豊	10	179.5	美浦石毛 532(-8)	芝	2500	2017		
11	9	4	6	6	ビュアソ 牡7	57	大野	9	76.6	美浦高橋 486(0)	芝	2500	2017		
12	0	1	3	3	シュベル 牡4	56	北村宏	1	3.4	美浦木村 510(+22)	芝	2500	2017		
13	1	2	8	10	サトノエ 牡4	55	吉田準	4	6.4	栗東角居 508(+6)	芝	2500	2017		
14	2	3	5	5	ステイバ 牡4	54	吉田豊	3	4.6	美浦尾関 434(+2)	芝	2500	2017		
15	3	4	4	4	ウインオ 牡6	51	石川	8	41.7	美浦鹿戸 440(-2)	芝	2500	2017		
16	4	4	8	9	ジョルジ 牡4	55	内田博	5	6.7	栗東鯨島 520(-6)	芝	2500	2017		
17	5	4	1	1	ヤマカツ 牡8	54	園分優	7	18.7	栗東大根 474(-6)	芝	2500	2017		
18	6	4	7	7	ウムブル 牡4	55	シュタ	2	3.6	美浦堀井 474(+14)	芝	2500	2017		
19	7	4	6	6	グラント 牡6	56	横山典	6	18.3	美浦武市 462(+14)	芝	2500	2017		
20	8	4	7	8	エアカー 牡6	54	大野	9	49.8	栗東藤田 482(-6)	芝	2500	2017		
21	9	4	2	2	マシェリ 牡4	50	菅原	10	107.3	美浦中川 482(-10)	芝	2500	2017		

## ▶ 予測レース

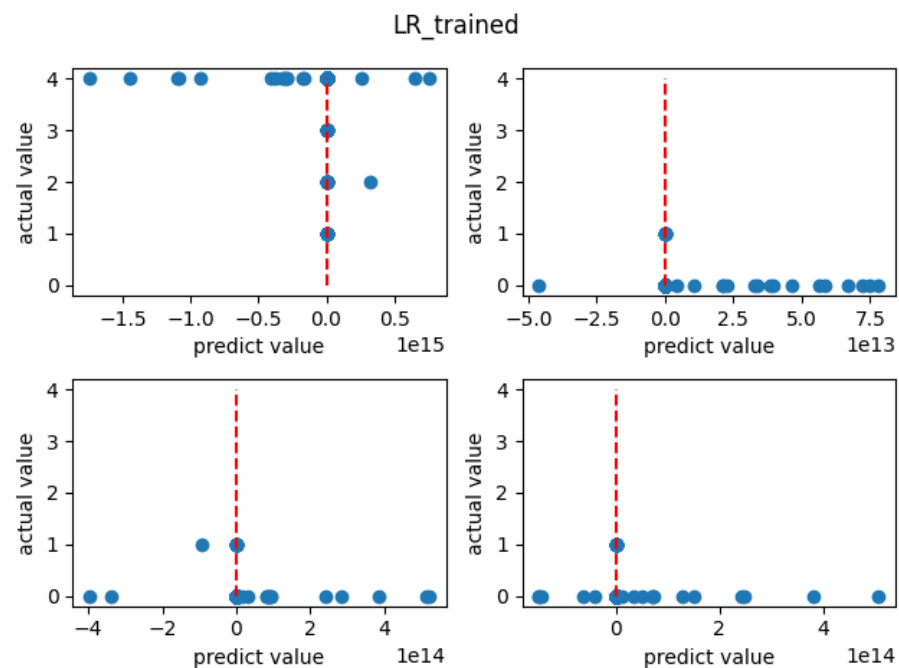
▶ 2021年12月26日、有馬記念

1	A	B	C	D	E	F	G	H	I	J	K
1	frame	number	name	sex_and_impost	jockey	popularity	odds	stable	weight		
2	0	1	1	ベルシアン 牡7	57	Cデム	11	62.3	栗東池江 500(-2)		
3	1	1	2	パンサラ 牡4	57	菱田	8	45.5	栗東矢作 480(+6)		
4	2	2	3	モズベッ 牡5	57	池添	14	146.2	栗東森田 488(0)		
5	3	2	4	メロディ 牡4	55	岩田望	15	146.8	栗東森田 352(-2)		
6	4	3	5	ディーブ 牡4	57	和田竜	5	20.9	栗東大久保 502(0)		
7	5	3	6	ウインキ 牡4	55	丹内	12	70	美浦宗像 480(+10)		
8	6	4	7	クロナジ 牡5	55	ルメー	2	2.9	栗東斎藤 478(0)		
9	7	4	8	ユーキャン 牡6	57	藤岡佑	16	150.6	栗東友道 512(+6)		
10	8	5	9	ステラヴ 牡3	55	Mデム	3	7.9	栗東須貝 498(+4)		
11	9	5	10	エフフォー 牡3	55	横山武	1	2.1	美浦鹿戸 516(+2)		
12	10	6	11	アリストラ 牡4	57	武豊	9	47.7	栗東音無 474(-6)		
13	11	6	12	シャドウ 牡5	55	横山典	13	144.5	美浦斎藤 476(-2)		
14	12	7	13	アカイ 牡4	55	幸	6	31.8	栗東中竹 516(+2)		
15	13	7	14	アサマンノ 牡3	55	田辺	10	51.1	美浦手塚 496(+4)		
16	14	8	15	キセキ 牡7	57	松山	7	43.8	栗東辻野 516(+4)		
17	15	8	16	タイトル 牡3	55	横山和	4	10.2	美浦栗田 476(+12)		

# 結果 2

## ▶ 線形回帰 学習結果

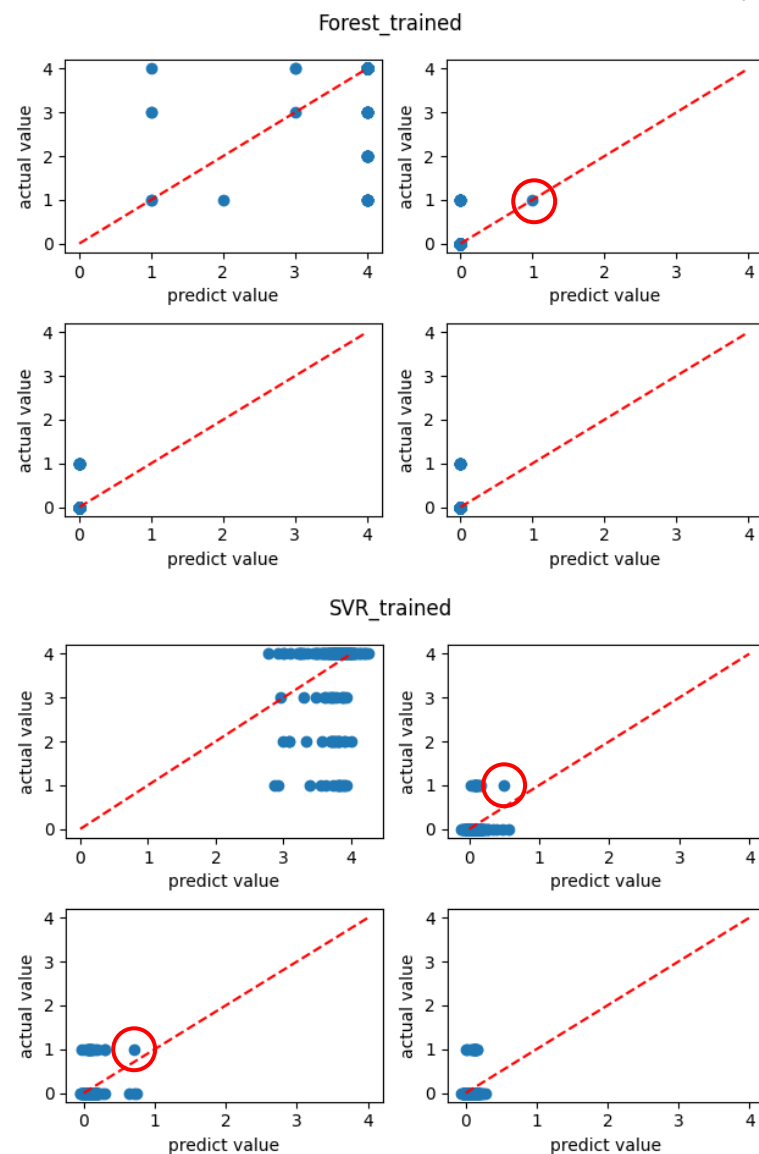
- ▶ 左上から 1～3着、1着、2着、3着
- ▶ 左上では 4着以下の馬は全て 4着として学習した
- ▶ 他は各着順なら 1、それ以外なら 0として学習した
- ▶ カス





# 結果 2

- ▶ ランダムフォレスト、SVR 学習結果
  - ▶ 前項と同様の方法で学習した
  - ▶ 概ねカスだが、比較的マシな結果になったものもあった



# 結果 2

## ▶ r2スコア

### ▶ 線形回帰

- ▶ -6.324835761494578
- ▶ -3.022821838984581
- ▶ -7.463996471215175
- ▶ -4.994346970870851

### ▶ SVR

- ▶ -0.025249379440665
- ▶ -0.038155320959743
- ▶ -0.042899213308601
- ▶ -0.004615312368619

### ▶ ランダムフォレスト

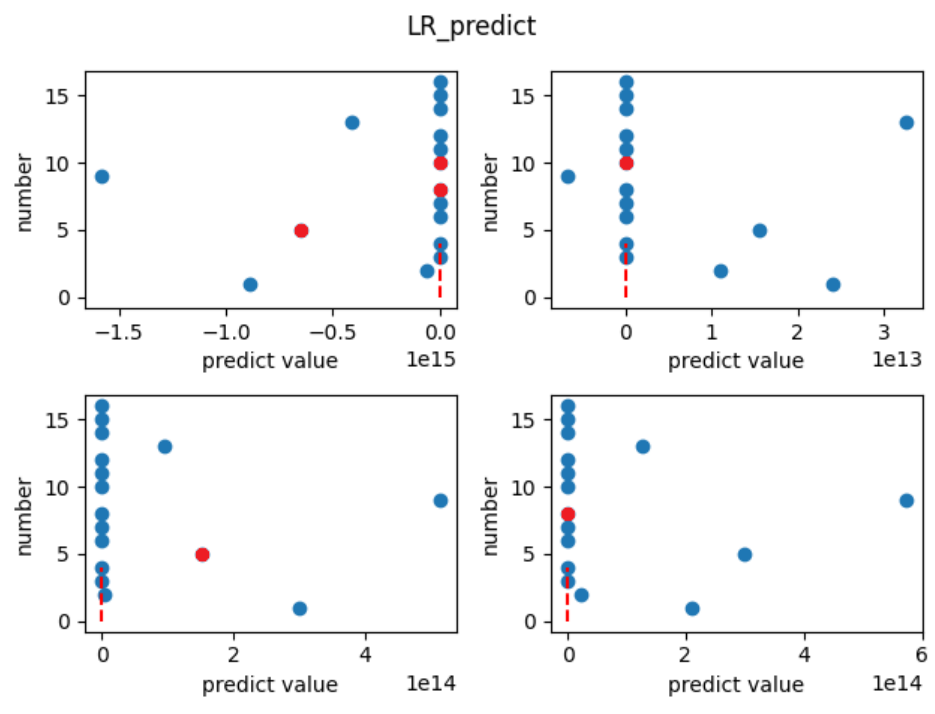
- ▶ -0.209251101321586
- ▶ -0.005494505494505
- ▶ -0.102409638554216
- ▶ -0.082840236686390

- ▶ 全ての方法において0を下回っている
- ▶ 線形回帰が最もスコアが低く、離れてランダムフォレスト、僅差でSVRが最もいいスコアとなっている

# 結果 3

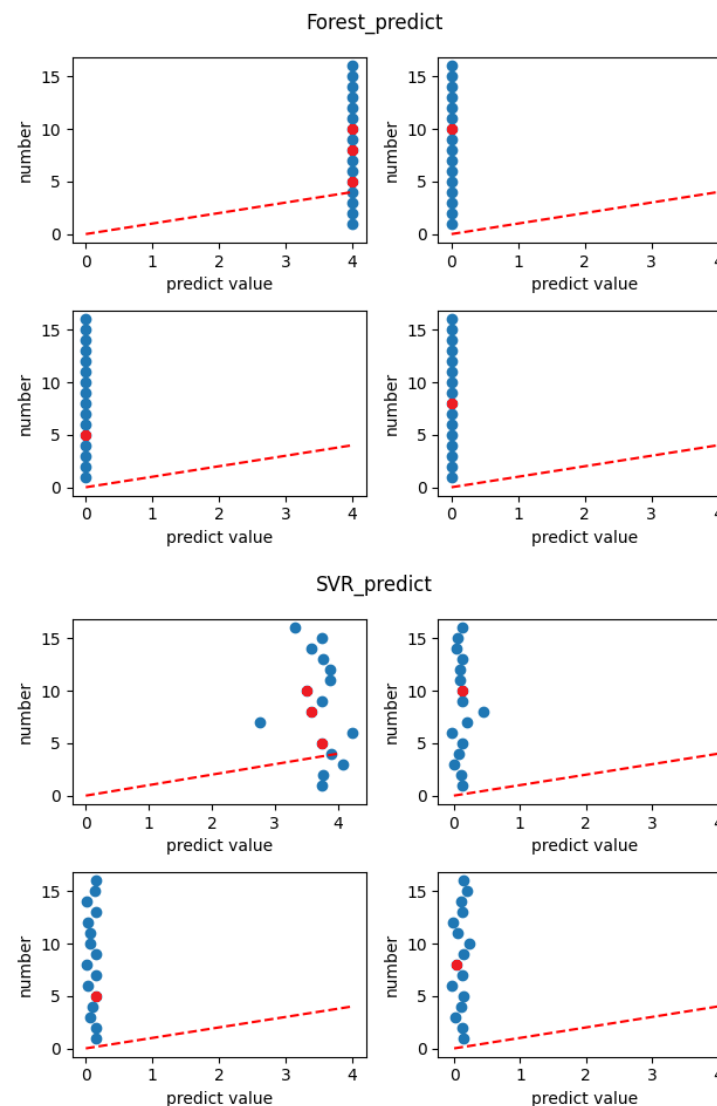
## ▶ 線形回帰 予測結果

- ▶ 縦軸は馬の番号、赤丸で示したのが実際の1～3着の馬である
- ▶ 左上において、結果2と同様に、1～4の整数値で学習したのにも関わらず0以下の結果が返ってきた



# 結果 3

- ▶ ランダムフォレスト、SVR 予測結果
  - ▶ ランダムフォレストでは、すべての結果が0となった
  - ▶ SVRでは、左上において実際の1～3着が比較的高い評価であった
  - ▶ また、5番（実際の2着）が2着になる確率が比較的高いと予測された



# 考察

- ▶ 今回の試行のほかに、中山芝2000mと2022年中山金杯、東京芝2000mと2021年天皇賞秋でも同様の予測を行ったが、今回の結果とほぼ同様の結果となった
- ▶ 1回の競馬におけるデータが少数かつ、それまでの馬の成績を鑑みずに予測を行ったため、正確な予測が難しかったと考える。
- ▶ また、要素の重みづけを行わずに正規化してしまったため、重要度の低い要素と高い要素が同じ扱いになってしまったことも挙げられる

# 今後の展望

- ▶ 過去の馬の成績も集め、より精度の高いものにしていきたい
- ▶ また、その際に膨大な量のスクレイピングを必要とするため、なるべくサーバに負荷のかからない方法を模索したい
- ▶ 今回は買い目の計算まで到達しなかったが、いずれ予測精度が高まれば実装したい

# 感想

- ▶ 想定していたよりも学習がうまくいかず、中途半端な結果になってしまった
- ▶ 高熱の中スライドを作ったので、しんどかった
- ▶ というかnetkeiba公式のCP予想でも回収率は100%に届かないのに素人の作ったモデルで100%は無理があったのでは？

ちなみに有馬記念は三連単当てました

# 参考文献

- ▶ GitHub/rytar/keiba-ml-preprocessing-tool
  - ▶ <https://github.com/rytar/keiba-ml-preprocessing-tools>
- ▶ Qiita/DS74/機械学習（回帰モデル）をscikit-learnを用いて様々な手法を実践してみた
  - ▶ <https://qiita.com/DS27/items/aa3f6d0f03a8053e5810>
- ▶ Pythonでデータサイエンス/scikit-learnで回帰モデルの結果を評価する
  - ▶ <https://qiita.com/DS27/items/aa3f6d0f03a8053e5810>