

Keiba Predict Diaries

Nov. 2019



そもそも、何で競馬予測をやってるの？

競馬予測を始めた経緯 / 目的:

- データ分析を通じたドメイン知識の獲得【◎】
 - 素人であるほど、先入観無く始められるメリットも
 - この発想はまさに『データドリブン思考』(の一例)
- 機械学習による予測精度の検証【○】
 - 競馬データは、機械学習でPDCAサイクルを回すのに好適
 - 手っ取り早く、予想を出した上でレースを観戦したい
 - 知名度の高いG1レースはじっくり考えられる(考えたい)がG2, G3レースまでは手が回らない...
- 金儲け【×】
 - 決して、金儲け目的ではありません...



競馬はスポーツorギャンブル？

競馬とは？：

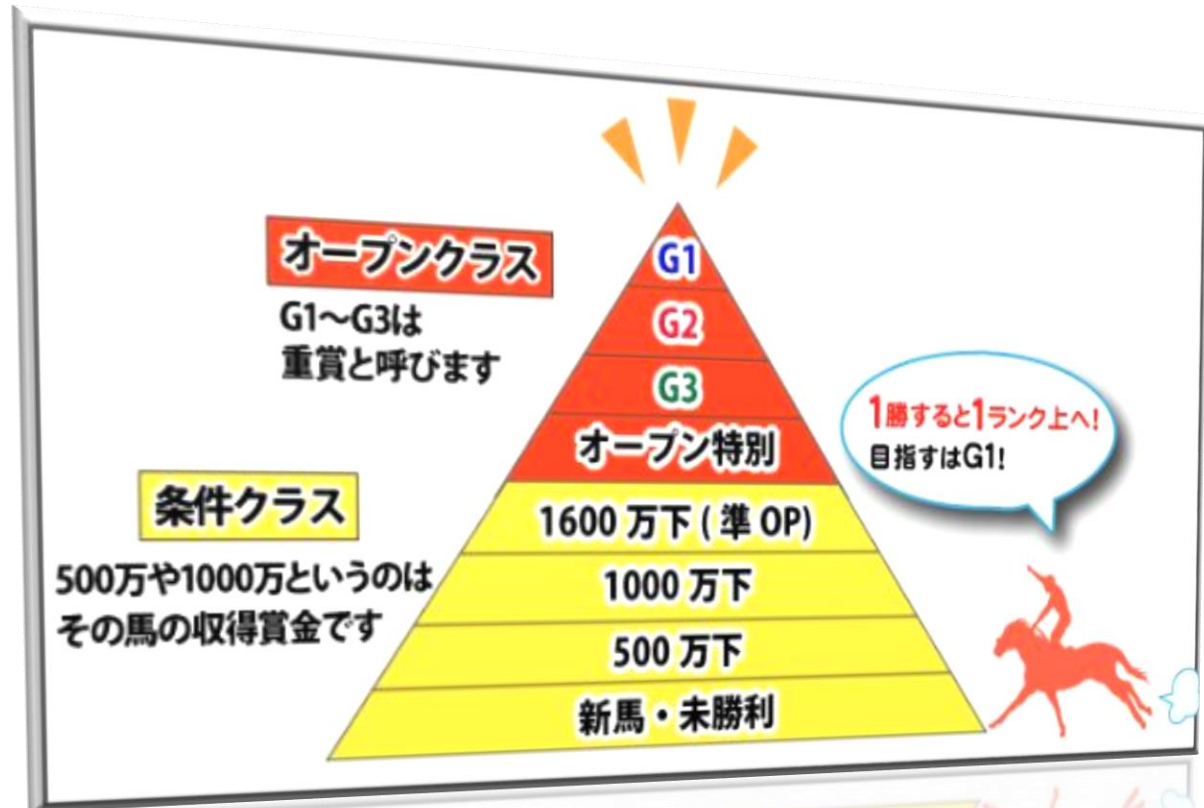
競馬とは、騎手が馬(サラブレッド)に乗り、決められたコースの決められた距離を決められた方向で周回し着順を競う競技のこと

- ・ コース： 芝 / ダート / 障害の3種類
- ・ 競馬主催者：
 - ・ **中央競馬(JRA)** / 地方競馬の2種類
 - ・ 中央競馬の競馬場は全国に10ヵ所
{ 01:札幌, 02:函館, 03:福島, 04:新潟, 05:東京, 06:中山, 07:中京, 08:京都, 09:阪神, 10:小倉 }
- ・ 馬券の正式名称：
『勝馬投票券(かちうまとうひょうけん)』



中央競馬はG1ランクを頂点とした階級制 レース体系:

G1～G3レースを重賞レースと呼ぶ



ご存知の通り、様々な賭け方がある 馬券の種類:

No	種類	説明
1	単勝	1着になる馬を当てる馬券
2	複勝	3着までに入る馬を当てる馬券
3	応援馬券	1頭の馬の「単勝」と「複勝」を同時に購入できる馬券
4	枠連	1着と2着になる馬の枠番号の組合せを当てる馬券
5	馬連	1着と2着になる馬の馬番号の組合せを当てる馬券
6	馬単	1着と2着になる馬の馬番号を着順通りに当てる馬券
7	ワイド	3着までに入る2頭の組合せを馬番号で当てる馬券
8	3連複	1着、2着、3着となる馬の組合せを馬番号で当てる馬券
9	3連単	1着、2着、3着となる馬の馬番号を着順通りに当てる馬券
10	WIN5	JRAが指定する5レース全ての1着馬を当てる馬券

No	種類	払い戻し例 (※100円で1枚購入した場合)					
		2019武蔵野S(11/9(土))			2019エリザベス女王杯(11/10(日))		
1	単勝	u07	2,520円	9番人気	u02	540円	3番人気
2	複勝	u07	690円	11番人気	u02	190円	3番人気
		u09	580円	8番人気	u06	330円	5番人気
		u15	2,110円	13番人気	u11	140円	2番人気
3	応援馬券	-	-	-	-	-	-
4	枠連	w04-w05	3,890円	17番人気	w01-w03	2,600円	10番人気
5	馬連	u07-u09	21,070円	56番人気	u02-u06	3,380円	11番人気
6	馬単	u07→u09	45,590円	113番人気	u02→u06	5,440円	18番人気
7	ワイド	u07-u09	5,600円	54番人気	u02-u06	970円	11番人気
		u07-u15	16,360円	76番人気	u02-u11	430円	3番人気
		u09-u15	24,840円	80番人気	u06-u11	940円	9番人気
8	3連複	u07-u09-u15	422,390円	319番人気	u02-u06-u11	4,060円	9番人気
9	3連単	u07→u09→u15	2,353,630円	1,795人気	u02→u06→u11	26,480円	72番人気
10	WIN5	-	-	-	-	-	-



今回は3連複を予測ターゲットとする

予測ターゲットの選定:

- ドメイン知識の獲得を踏まえると、可能な限り幅広い賭け方を選択するのが望ましい
- 3連単は夢があるが、現実的ではない
- 3連複なら後述の人気-着順のパターン分類で当てられるかも？(淡い期待)

三連複【さんれんぷく】

難易度 ★★★★★ 配当 ★★★★★

着順は関係なく1着から3着にくる3頭を当てる馬券種です。
着順上位3頭を選んでいればどの馬が1着でどの馬が2着でも的中となります。

1着 2着 3着

ゴール

8 13 4

馬券イメージ

2012年06月04日 第12レース

3連複 100円

4-8-13

左図での的中買い目

4-8-13



人気と着順のパターンに着目してみる

着順パターンの分類に関するクイズ:

【Q1.】 人気1位の馬が1着となる確率は0%

[人気]: 1|*|*|... → [着順]: 1|*|*|...

**【Q2.】 1-3番人気の少なくとも一頭の馬が
1-3着内に入る確率は0%**

[人気]: 1|2|3|... → [着順]: *|1|*|... or 2|*|*|... or
3|1|*|... or 1|3|2 or 他多数

**【Q3.】 1-3番人気の馬が
全て1-3着圏外となる確率は0%**

[人気]: 1|2|3|... → [着順]: 8|4|15|... or 7|5|12|... or 他多数



人気と着順のパターンに着目してみる

着順パターンの分類に関するクイズの回答:

【A1.】 人気1位の馬が1着となる確率は30%

[人気]: 1|*|*|... → [着順]: 1|*|*|...

【A2.】 1-3番人気の少なくとも一頭の馬が
1-3着内に入る確率は90%

[人気]: 1|2|3|... → [着順]: *|1|*|... or 2|*|*|... or
3|1|*|... or 1|3|2 or 他多数

【A3.】 1-3番人気の馬が
全て1-3着圏外となる確率は10%

[人気]: 1|2|3|... → [着順]: 8|4|15|... or 7|5|12|... or 他多数



人気1位の馬が1着となる確率は30%

着順パターンの分類:

- 重賞レース全体で **30%** ($=396 / 1,303$)
- レースランク別ではG1が35%と若干高い
G1のみ: **35%** ($=81 / 232$)
G2のみ: **33%** ($=121 / 372$)
G3のみ: **28%** ($=194 / 699$)

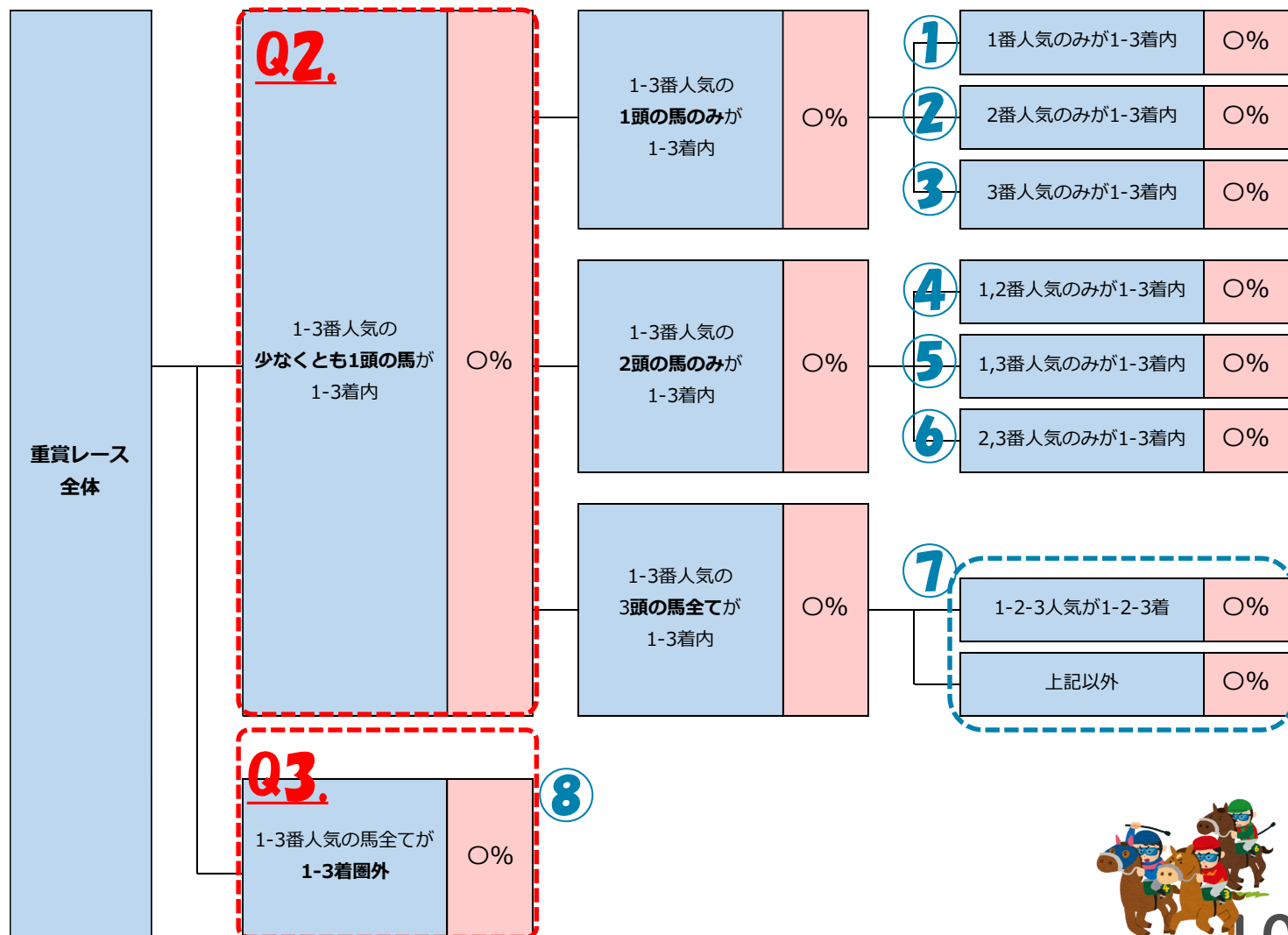
◆ 人気1位の馬が1着となる確率 (直近10年)

No	年	G1			G2			G3			重賞(G1~G3)		
		該当数	レース数	確率	該当数	レース数	確率	該当数	レース数	確率	該当数	レース数	確率
1	2010	11	22	50%	9	37	24%	14	67	21%	34	126	27%
2	2011	8	21	38%	16	37	43%	17	65	26%	41	123	33%
3	2012	6	24	25%	11	38	29%	17	71	24%	34	133	26%
4	2013	9	24	38%	12	38	32%	22	71	31%	43	133	32%
5	2014	6	24	25%	10	39	26%	21	72	29%	37	135	27%
6	2015	7	24	29%	8	39	21%	13	71	18%	28	134	21%
7	2016	9	24	38%	14	39	36%	23	74	31%	46	137	34%
8	2017	11	26	42%	15	37	41%	22	75	29%	48	138	35%
9	2018	9	29	31%	14	38	37%	27	73	37%	50	140	36%
10	2019	5	14	36%	12	30	40%	18	60	30%	35	104	34%
-	計	81	232	35%	121	372	33%	194	699	28%	396	1,303	30%



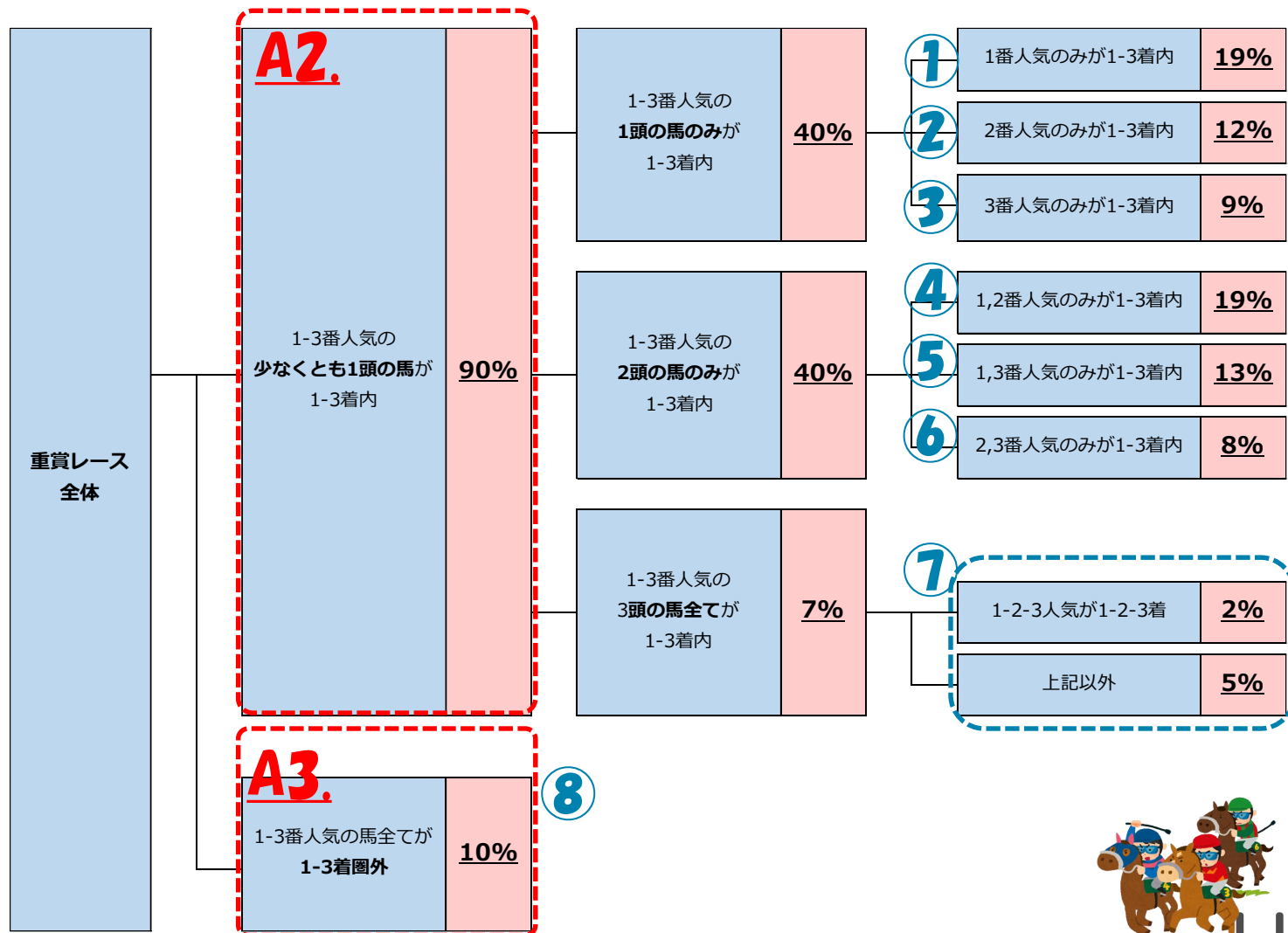
1,2,3番人気の着順を①-⑧の8パターンに分類

着順パターンの分類:



1,2,3番人気の着順を①-⑧の8パターンに分類

着順パターンの分類:



1,2,3番人気の着順を①-⑧の8パターンに分類 着順パターンの分類:

11/9(土)~10(日)の重賞5レースの1-3着馬の人気

- ・ 武蔵野ステークス(G3): 9-8-13番人気(荒れるレース) (⑧)
- ・ 京都ジャンプステークス(G3): ①-5-③番人気 (⑤)
- ・ テイラー杯2歳ステークス(G2): ③-①-①番人気 (⑤)
- ・ エリザベス女王杯(G1): ③-①-①番人気 (⑤)
- ・ 福島記念(G3): ①-6-②番人気 (④)

⇒ 『⑤1,3番人気の2頭の馬のみが1-3着内』
のレースが5レース中3レース(60%)!!

⇒ 着順パターンが予測できれば、3連複の確度が上がる
・ ラスト1pieceを埋めればよい(網羅賭けでもよい)



予測ツールを自動作成し活用中

Keiba-Predictor-Tools (略して KPT):

重賞レースが開催される週末の前日までに
自動でレポートを作成・更新

< 予測ツール機能 >

- ・ 0. 重賞レース出馬表
- ・ 1. 単勝予測結果
- ・ 2. 過去の不整合率の推移
- ・ 3. 1-3番人気馬の着順分類予測
- ・ 4. 3連複予測結果
- ・ 5. オッズの変動状況



馬柱形式の出馬表をスクレイピング

0. 重賞レース出馬表:

11R

天皇賞 (秋)

GI

4回 9日目 芝2000m(左) 15:40発走

8		7		6		5		4		3		2		1		枠番
16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	馬番
ドバイマジエスティ アルアイン	ディーブインバクト ウインブライト サマーエタニティ	ステイゴール ミスタンコール ウグネリアン	ディーブインバクト ラストグルーヴ ランフォザローゼス	ディアデラノビア ドレッドノータス	ハービンジャー ラグジャリー ゴーフオザサミツト	ロードカナロア サートウルナリア	ディーブインバクト ダノンプレミアム インディアナギャル	ウイキウィキ マカヒキ	ディーブインバクト シリアスアティテュード スティッフエリオ	キングカメハメハ ムービーインディゴ ユーキヤンスマイル	クロフネ アエロリット アステリックス	ピラミマ スワーヴリチャード	ハーツクライ ケイアイイーガーベラ ケイアイノーテック	ディーブインバクト フサイチパンドラ アーモンドアイ	ロードカナロア フレッチャリヴィエラ カデナ	母馬名
▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	母馬名
鹿毛	鹿毛	芦毛	鹿毛	鹿毛	栗毛	黒鹿毛	青鹿毛	鹿毛	鹿毛	鹿毛	鹿毛	芦毛	栗毛	鹿毛	鹿毛	脚質
58.0 牡5	58.0 牡5	58.0 牡4	56.0 牡3	58.0 セ6	58.0 牡4	56.0 牡3	58.0 牡4	58.0 牡6	58.0 牡5	58.0 牡4	56.0 牝5	58.0 牡5	58.0 牡4	56.0 牝4	58.0 牡5	斤量 馬齢
北村友	松岡	福永	Mデム	坂井	北村宏	スミヨ	川田	武豊	丸山	岩田康	戸崎圭	横山典	幸	ルメー	藤岡佑	騎手
栗東 池江	美浦 畠山	栗東 友道	美浦 藤沢和	栗東 矢作	美浦 藤沢和	栗東 角居	栗東 中内田	栗東 友道	栗東 音無	栗東 友道	美浦 菊沢	栗東 庄野	栗東 平田	美浦 国枝	栗東 中竹	調教師
54.6 10人氣	87.9 12人氣	13.2 4人氣	46.3 9人氣	135.0 14人氣	298.0 16人氣	3.8 2人氣	12.0 3人氣	28.6 8人氣	126.1 13人氣	24.3 6人氣	26.2 7人氣	19.1 5人氣	163.5 15人氣	1.7 1人氣	84.5 11人氣	馬体重 オッズ
... V	... V	... V	... V	... V	... V	... V	... V	... V	... V	... V	... V	... V	... V	... V	... V	MY予想



3種類の機械学習モデルの予測を出力

1. 単勝予測結果:

【天皇賞(秋)(G1) 2019/10/27(日)】					【天皇賞(秋)(G1) 2019/10/27(日)】					ソートもできます			
【天皇賞(秋)(G1)】 2019/10/27(日), 05_東京, 芝左2000m					【天皇賞(秋)(G1)】 2019/10/27(日), 05_東京, 芝左2000m					①レースタイム予測		②1-3着内予測	
Show 20 entries					Show 20 entries					③ダークホース予測			
枠番	馬番	馬名	騎手	調教師	人気	オッズ	予測タイム(sec)	1_順位	1-3着内確率(%)	2_順位	ダークホース確率(%)	3_順位	
		All	All	All									
1	1	カデナ	藤岡佑介	中竹和也	1	1.7	116.1	1.5	64.9	1	0.8	15	
1	2	アーモンドアイ	C. ルメール	国枝栄	2	3.5	116.1	1.5	37.7	5	1.2	10	
2	3	ケイアイノーテック	幸英明	平田修	3	9.0	124.9	9.0	49.2	2	0.7	16	
2	4	スワーヴリチャード	横山典弘	庄野靖志	4	9.7	125.5	10.0	46.3	3	0.8	14	
3	5	アエロリット	戸崎圭太	菊沢隆徳	5	11.5	125.7	12.0	33.9	6	0.8	13	
3	6	ユーキャンスマイル	岩田康誠	友道康夫	6	15.2	122.3	5.0	29.1	8	0.8	12	
4	7	スティッフフェリオ	丸山元気	音無秀孝	7	25.8	125.6	11.0	31.0	7	1.4	9	
4	8	マカヒキ	武豊	友道康夫	8	29.6	116.5	3.0	44.0	4	1.1	11	
					9	90.9	125.9	15.0	16.6	9	51.2	1	



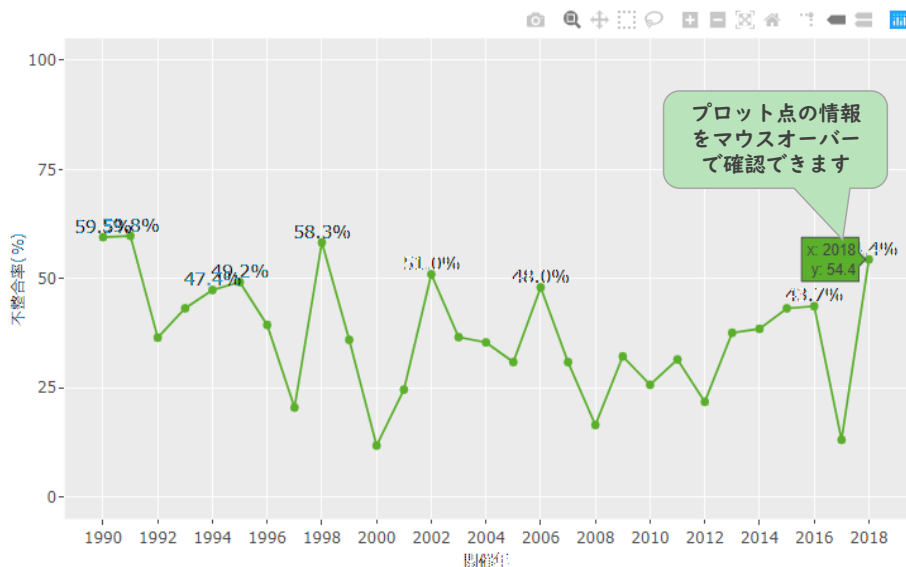
人気と着順の不整合率の推移を確認可能

2. 過去の不整合率(人気-着順)の推移:

[アルデミスステークス_2012y_2018y]

[スワンステークス_1990y_2018y]

[天皇賞(秋)_1990y_2018y]



人気と着順の乖離が大きいほど
不整合率(0-100%)は大きくなる

[スワンステークス_1990y_2018y]

[天皇賞(秋)_1990y_2018y]

Search:

開催年	不整合率(%)	頭数	人気別着順(1-3位)	人気別着順(全体)	【着順】 2番人気-4番人気 -6番人気----
2018	54.4	12頭	2-4-6	2 4 6 5 8 10 3 7 9 1 12 11	
2017	13.1	18頭	1-2-13	1 2 13 3 9 4 15 12 6 10 7 14 8 18 11 17 16 5	
2016	43.7	15頭	1-7-6	1 7 6 4 9 11 3 12 5 10 14 2 15 8 13	
2015	43.2	18頭	1-10-6	1 10 6 5 4 17 15 3 2 8 16 9 13 18 12 11 7 14	
2014	38.5	18頭	5-2-1	5 2 1 16 10 4 6 11 9 12 13 8 7 3 14 15 17 18	
2013	37.6	17頭	5-1-3	5 1 3 8 4 13 11 7 15 2 10 9 12 16 17 6 14	
2012	21.8	18頭	5-1-2	5 1 2 4 3 8 16 10 6 15 13 12 7 17 14 18 9 11	
2011	31.5	17頭	7-2-6	7 2 6 1 10 3 13 8 9 4 17 16 18 5 11 12 14	
2010	25.7	18頭	1-4-2	1 4 2 17 11 3 16 13 10 15 8 14 9 6 18 12 7 5	



1-3番人気馬の着順パターンを8分類し 分類パターンを予測

3. 1-3番人気馬の着順分類予測:

アルテミスステークス スワンステークス 天皇賞(秋)

Show 8 entries

Search:

分類	凡例	発生確率 (%)	当該レース	当該レースランク	直近1年間
All					
1: 1番人気のみ (1_pop1)	1 * *	20.3	0.84	0.99	1.01
2: 2番人気のみ (2_pop2)	2 * *	9.5	0.69	0.86	0.89
3: 3番人気のみ (3_pop3)	3 * *	9.6	1.39	1.35	1.92
4: 1,2番人気のみ (4_pop12)	1 2 *	19.8	0.96	1.02	0.84
5: 1,3番人気のみ (5_pop13)	1 3 *	13.8	2.00	0.99	1.21
6: 2,3番人気のみ (6_pop23)	2 3 *	8.2	2.41	1.11	1.44
7: 1,2,3番人気すべて着内 (7_pop123)	1 2 3	6.5	0.94	0.75	0.76
8: 1,2,3番人気すべて着外 (8_darkhorse)	* * *	12.4	0.72	1.05	0.83

レース毎に
8分類の発生確率
を予測

8分類の発生確率(水色セル)は、
水準差がある為、ベース統計量から
のリフト値(3種類の緑セル)も
参考情報として活用

・[当該レース]: 過去の天皇
賞(秋)だけで集計した発生確率
に対するリフト値

・[当該レースランク]: 過去
の全GIレースのみで集計した
発生確率に対するリフト値

・[直近1年間]: 前年1年間の
全重賞レースのみで集計した
発生確率に対するリフト値



3連複の全組み合わせを抽出し、 単勝予測結果の平均でスコアリング

4. 3連複予測結果:

【アルテミスステークス(G3) 2019/10/27(日)] 【スワンステークス(G2) 2019/10/27(日)]

【天皇賞(秋)(G1) 2019/10/27(日)]

Show entries Search:

発生パターン毎に フィルタリングして利用	分類	馬番	人気	平均予測 タイム (sec)	1-3着内確 率(%)	ダークホース 確率(%)
5:						
5: 1,3番人気のみ (5_pop13)		2 9 14	1 3 4	122.17	14.78	0.00
5: 1,3番人気のみ (5_pop13)		2 9 16	1 3 8	119.17	14.04	0.00
5: 1,3番人気のみ (5_pop13)		2 9 4	1 3 5	122.23	10.82	0.00
5: 1,3番人気のみ (5_pop13)		2 9 6	1 3 7	122.20	9.90	0.00
5: 1,3番人気のみ (5_pop13)		2 9 5	1 3 6	121.10	9.29	0.00
5: 1,3番人気のみ (5_pop13)		2 9 8	1 3 9	122.30	5.28	0.00
5: 1,3番人気のみ (5_pop13)		2 9 13	1 3 13	121.63	4.14	0.00

Showing 1 to 10 of 13 entries (filtered from 560 total entries)

us 1 2 Next

単勝予測結果を元に
3連複パターンの予測スコアを算出

- ・ [平均予測タイム] :
単勝予測結果の3頭のタイムの単純平均
- ・ [1-3着内確率(%)] [ダークホース確率(%)] :
単勝予測結果の3頭の予測確率の単純積

人気の観点で3連複パターンの
全組み合わせを計算



1時間単位でオッズ(予想→中間→最終)の変動状況をモニタリング

5. オッズの変動状況:

[アルテミスステーキス]

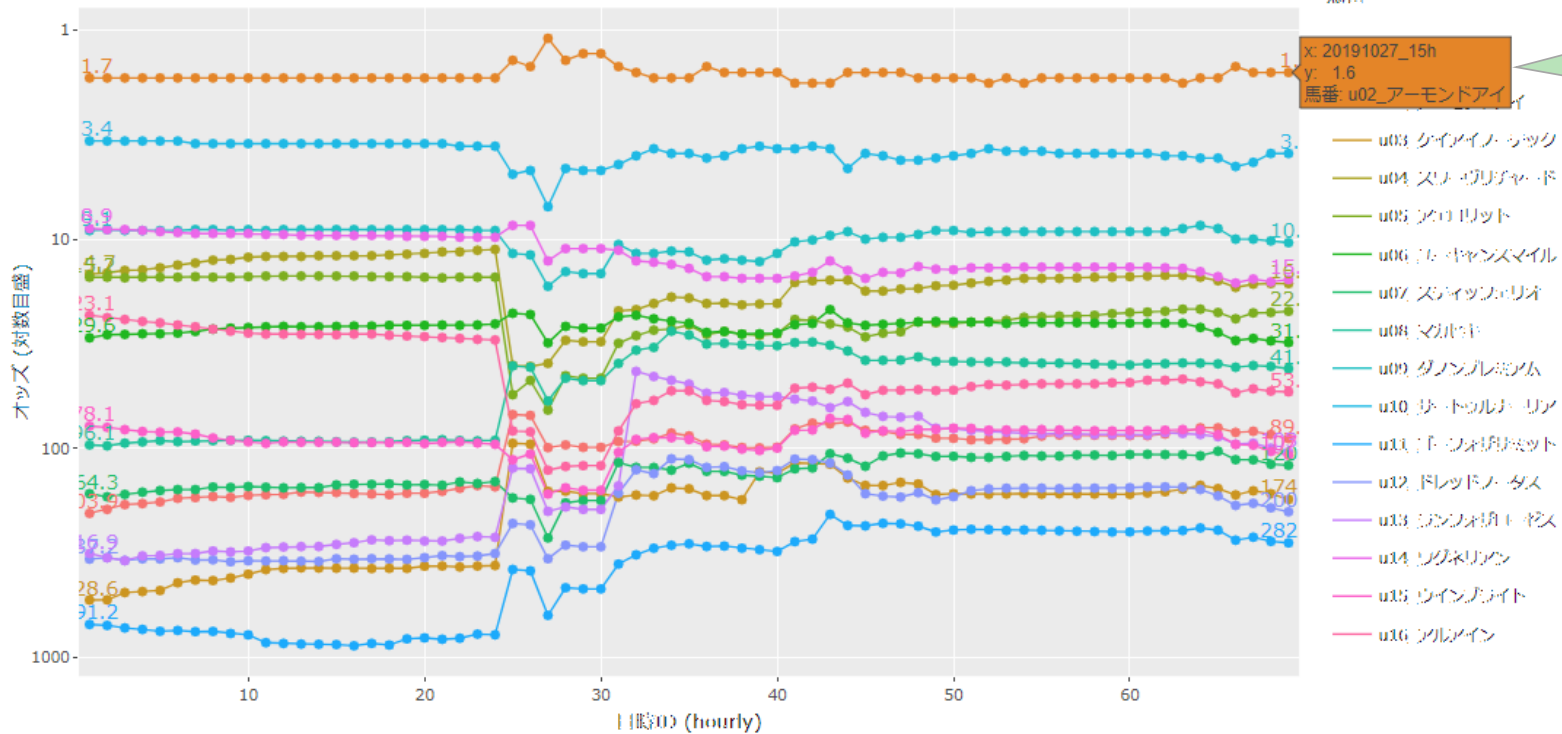
[スワンステーキス]

[天皇賞(秋)]

観望期間: [20191024_15h] ~ [20191027_15h]



馬番



プロット点の情報をマウスオーバーで確認できます



どの時点のオッズを使って予想すべきか？

オッズ情報の利用時点の見極め：

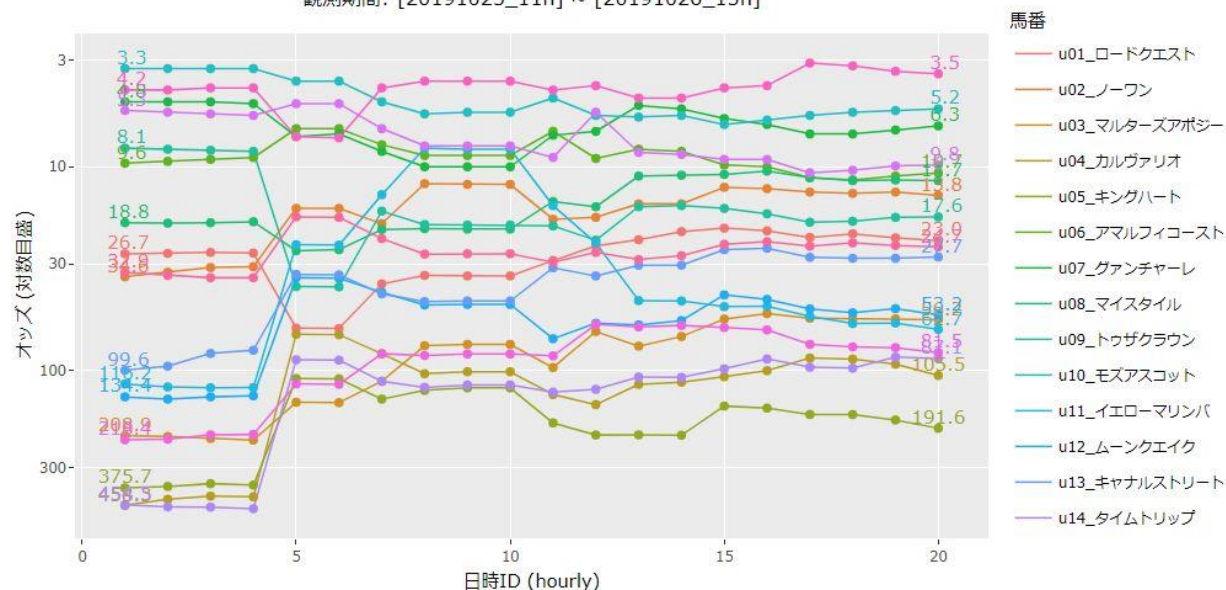
予想オッズ → (発売開始) → 中間オッズ → 最終オッズ

アルテミスステークス

スワンステークス

天皇賞(秋)

観測期間: [20191025_11h] ~ [20191026_15h]



中間オッズ以降で1-3番人気がほぼ確定した
段階で予測・購入するのが望ましいか？



データ分析で得られた知見をもとに 『競馬予測Nヶ条』を作成予定 今後取り組みたいこと①:

◆ 競馬予測Nヶ条 (当たらぬものは当たらぬものです)

- 一、人気1位の馬が1着となる確率は30%
- 一、1-3番人気の少なくとも一頭が1-3着内に入る確率は90%
- 一、勝負所の馬番は素数(0番)を選ぶべし!
- 一、00な場合のレースは荒れやすい(番狂わせが起きやすい)
- 一、開催場所、開催時期(シーズン)に関するもの
- 一、レース場(芝/ダート, 馬場コンディション, 天候)に関するもの
- 一、...

(参考) 什の掟 — じゅうのおきて (ならぬことはならぬものです)
<https://nisshinkan.jp/about/juu>



各視点でのドメイン知識の深化

今後取り組みたいこと②:

- 位置取り(馬番・枠番):

馬番の素数の法則?

好位置にいながら着順が悪い、出遅れ癖

- 競走馬の調子(パドック):

馬体重、体重増減

騎手が乗ったら落ち着いた/暴れた

- 血統:

血統は予測にどの程度効くのか?、何世代まで見るべき?

- 騎手・調教師:

騎手の巧拙の影響度、陣営の戦略

騎手は一日に数レースを走っている(Tough Guy!)



その他、面白いことを試行予定

今後取り組みたいこと③:

- 巷の競馬予測AIとの比較:

- Alphalmpact(ウマいい馬券), ...
- 予測結果の差異傾向を比較してみたい

- レース観戦時の補助情報表示:

- 競馬レースの動画で人気or馬名がリアルタイム表示されながら観戦できたら便利！
- ディープラーニングで「ゼッケン(馬番)」or「馬」を認識するモデルを構築し、エッジ推論



