



チーム：ウマ天気

2017年11月11日

<メンバー>

二見	加代子	(キヤノンITソリューションズ株式会社)
根岸	颯	(キヤノンITソリューションズ株式会社)
赤池	奈奈	(個人会員)
大石	直人	(非会員)

<サポート>

吉田	光輝	(個人会員)
舟城	亮一	(個人会員)

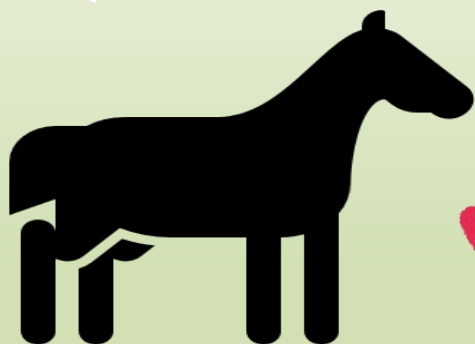
- ✧ 概要
- ✧ テーマを決めた目的と背景
- ✧ 活動実績
- ✧ データ集めとデータ選定
- ✧ 予想方法のアプローチ
- ✧ 方法1: 走行タイムを予想
- ✧ 方法2: 出走馬の属性から予想
- ✧ デモ
- ✧ みんなの感想

競馬も機械学習もシロウトだけど、
半年かけて勉強しながら、試行錯誤で

競馬予想にチャレンジしてきました

大テーマ: 動物と気象

動物といえば…ウマ!?



競馬予測が
いいんじゃない?

テーマを決めた目的と背景



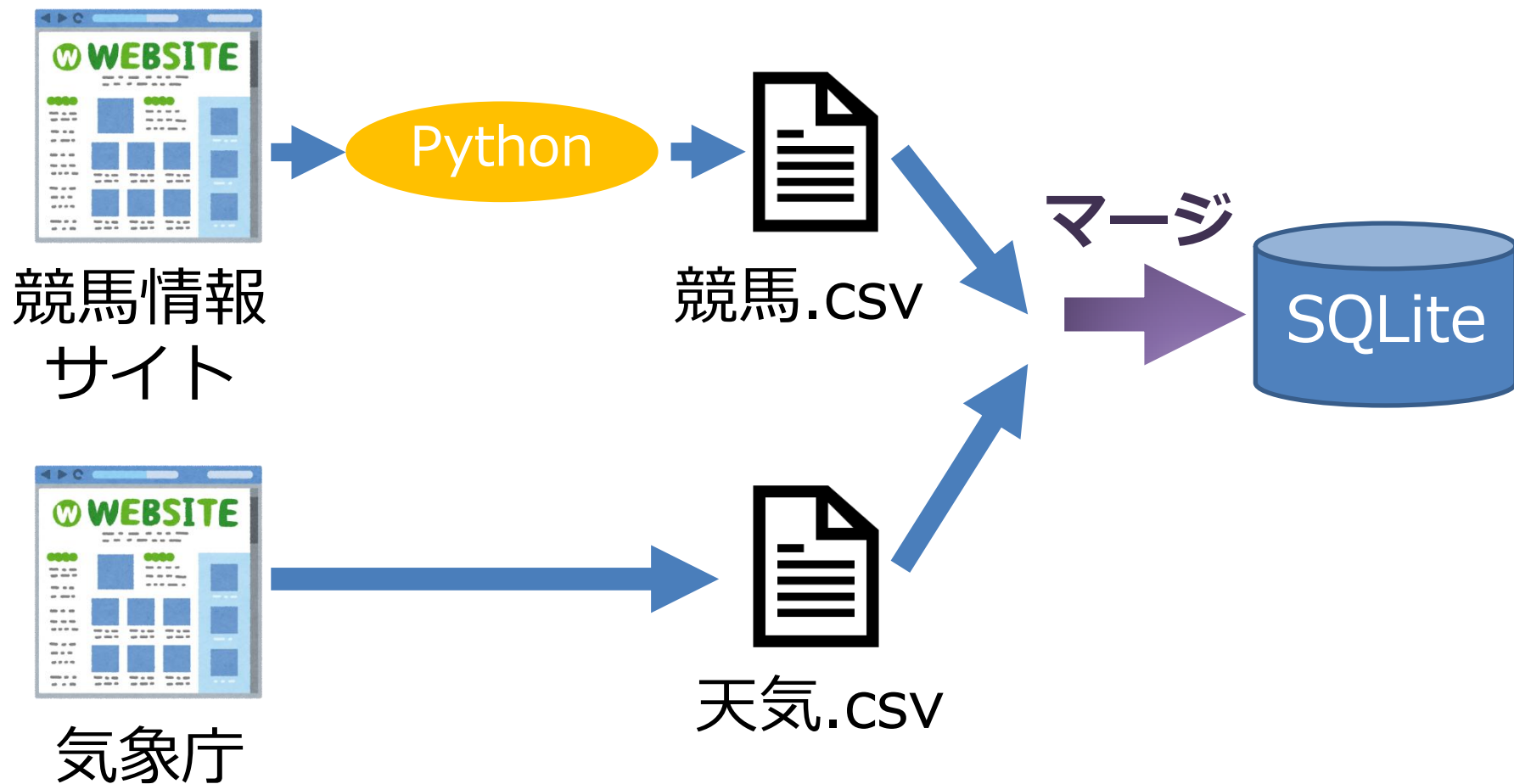
データが集めやすい

勉強しながらおカネもゲット
できたりして…



競馬予測に決定!





データ集めとデータ選定

ネット競馬

気象庁

データが扱いやすい

コストがかからない



データが足りないので…



Yahoo競馬

データ量が多い

やや扱いにくい



4・5月

- 分析内容を決定
- 競馬と気象（馬場状態、天気、湿度等）の相関を調査

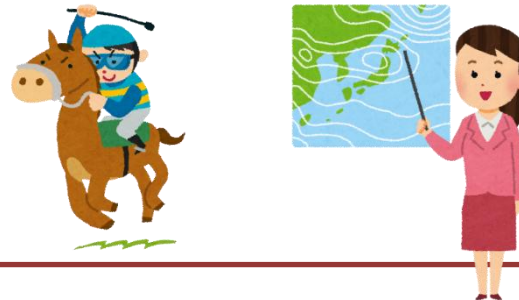


5・6月

- データ収集
- Rを使用してデータ分析・Microsoft Azureでデータマイニング
- 分析に有効なデータ属性を検討

7月以降

- データ分析



予想方法のアプローチ

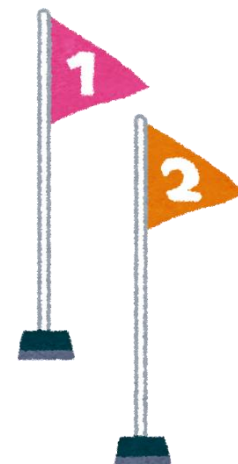
※ 方法1.

- ※ 出走馬の属性から「走行タイム」を予想する
- ※ タイム順で順位予想を得る
- ※ 機械学習としては「値予想」に該当する



※ 方法2.

- ※ 出走馬の属性から「順位」を予想する。
- ※ 1つの順位に複数の馬が存在しても良しとする
- ※ 機械学習としては「クラス分け」に該当する



方法1: 走行タイムを予想

※ 目的：走行タイムを予想する

※ 内容：

※ ネット競馬から2年分のデータを収集

※ Microsoft Azure Machine Learningで機械学習

※ 求める値を走行タイムとする

※ 結果：予想タイムと実績タイムの乖離が大きく
(1秒以上) 競馬の予想材料としては不適切

※ 考察：値の予想に誤差は必ず入るが、競馬は
「ハナの差」など僅差で争うため、誤差の影響
が大きい



方法2：出走馬の属性から予想

※ 目的：出走馬の順位を予想する

※ 内容：

※ 過去のレースの順位と、その時の出走馬、騎手、調教師の成績を属性として学習しておく

※ RのRandomForestライブラリを利用

※ 結果：

※ 当初簡単な属性（年間の勝利回数）で検証
→ 順位予想の確率が少し上がってきた

※ 本格的に属性を計算
→ ますます予想確率が高まった

※ 3位以内に入った時の「気温・湿度・風速・降雨量」を計算し、レース日の気象との差を学習
→ 順位予想に大きく影響

方法2：出走馬の属性から予想

※ 考察：

※ どんな属性を計算したらいいか？

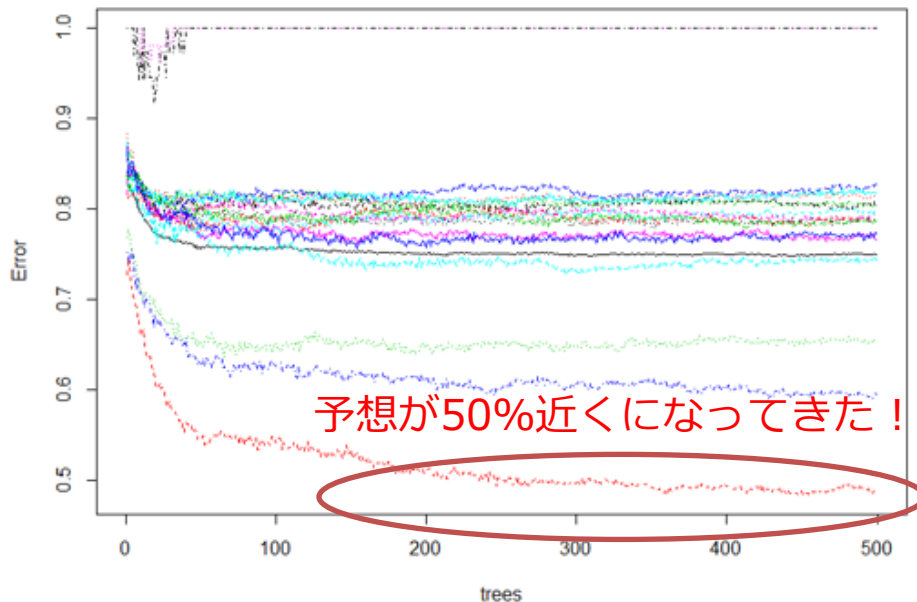
→ 人間が競馬予想することと同じかもしれない

※ 予想に最も影響を与える属性は「Odds」

→ 人間の知恵の結果を利用している？

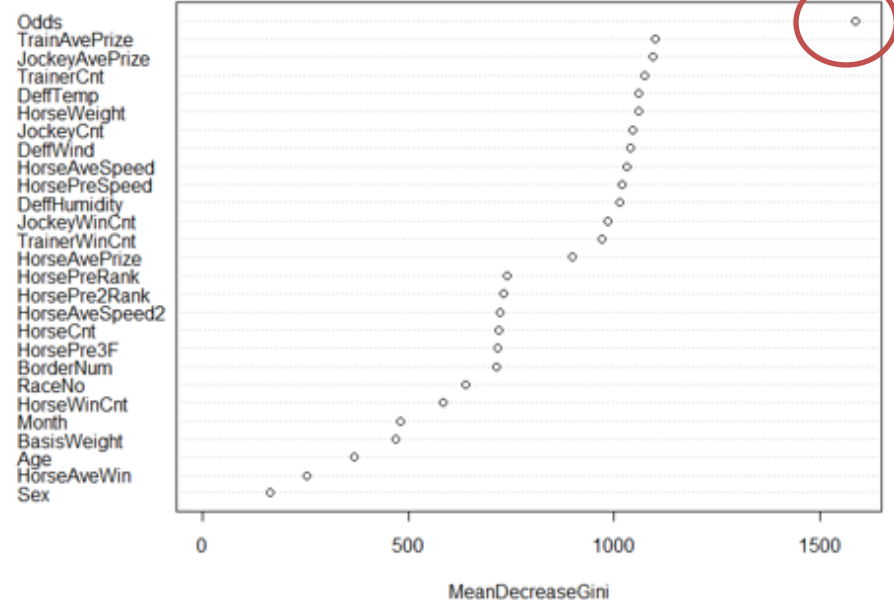
(万馬券は予想できない)

《予想のエラー率》



《予想に影響を与える属性》

Nakayama.rf



参考：計算した属性

- 過去4レースの平均速度 (HorseAveSpeed)
- 過去4レースの3位以内確率 (HorseAveWin)
- 馬の平均獲得賞金 (HorseAvePrize)
- 騎手の平均獲得賞金 (JockeyAvePrize)
- 前レースの順位 (HorsePreRank)
- 前々レースの順位 (HorsePre2Rank)
- 前レースの速度 (HorsePreSpeed)
- レース実施月 (Month)
- 調教師の平均獲得賞金 (TrainAvePrize)
- 前回の上がり3 F (HorsePre3F)
- 3位以内時の平均気温 とレース時平均気温の差 (DeffTemp)
- 3位以内時の平均湿度 とレース時平均湿度の差 (DeffHumidity)
- 3位以内時の平均風速 とレース時平均風速の差 (DeffWinWind)
- 過去4レースの平均速度(同一競馬場、コース) (HorseAveSpeed2)
- 馬の年間出場回数 (HorseCnt)
- 馬の年間3位以内回数 (HorseWinCnt)
- 騎手の年間出場回数 (JockeyCnt)
- 騎手の年間3位以内回数 (JockeyWinCnt)
- 調教師の年間出場回数 (TrainerCnt)
- 調教師の3位以内回数 (TrainerWinCnt)

みんなの感想

- ※ 根岸：R, Python, SQLiteなど、データ解析に必要な知識を身につけることができ、大変有意義な勉強会でした！
- ※ 二見：一人で勉強していたら挫折してしまったと思うが、みんなで楽しく学べたことが良かった！
この貴重な経験を生かして、機械学習をこれからも続けたい。
- ※ 赤池：競馬データはネット上に多く公開されていたが、加工しやすく形式が統一されたデータ探し・データ加工が手間どった。
分析に入る前のデータ準備が重要かつ時間がかかると感じた。
- ※ 舟城：はじめて機械学習に触れ、その実力に感動。課題解決にどう使うか？という知恵が重要と感じた。勉強しながらだったので、サポートなのに当事者になってしまったことが反省。
- ※ 大石：
- ※ 吉田：実データでの予想が結構難しく実体験ができ良かった。
レース等 外乱がある予想は難しい走行タイム予想をやったが誤差の影響が大きく良い結果が得られず残念。誤差を排除する努力を続けると使えるようになるか？？

最新情報は
こちらをご参照ください



<http://aitc.jp>



<https://www.facebook.com/aitc>



ハルミン
AITC非公式イメージキャラクター