



活動実績

和知 太亮

# 経歴

- ・出身は茨城県
- ・令和1年 茨城県日立市立日立第一高等学校 卒業
- ・令和1年 会津大学 入学
- ・小学1年生から現在までサッカーを続けている。

# 製作物

- ・ API  
Line Botの作成
- ・ スクレイピング  
loto7の当選番号の予測  
自動画像収集
- ・ データの操作  
(pandas, numpy, scikit-learn)  
kaggleでのデータの扱い



[Dio] 名言94.不死身・不老不死・スタンドパワー

午前 11:47



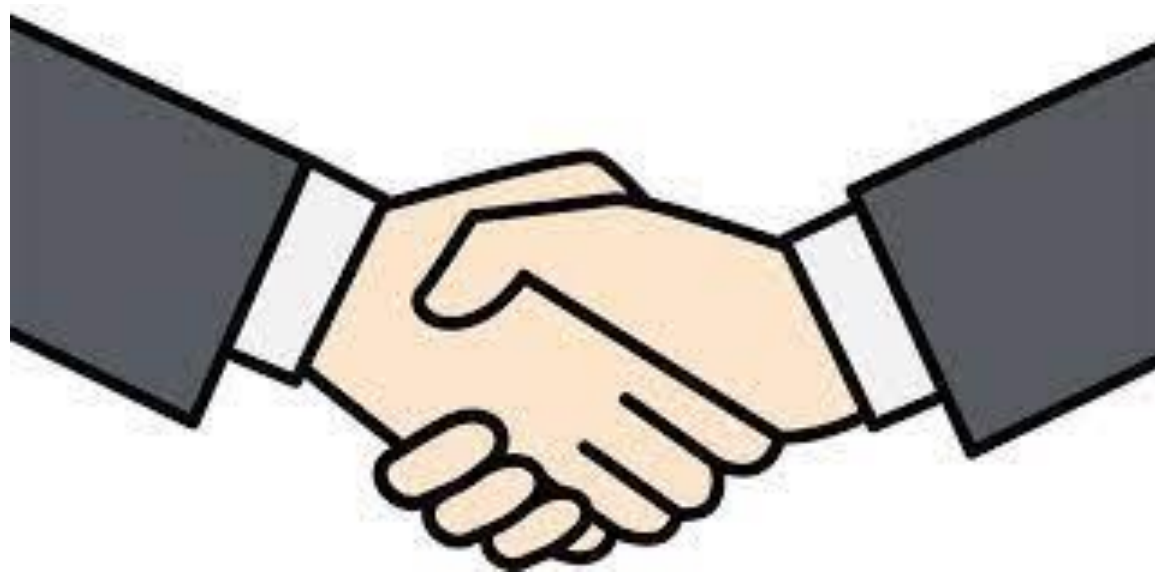
午前 11:47

保存 | 名前を付けて保存 | 転送 | タイムライン | Keep

例 [Line bot]

## 製作物の紹介(2つ)

- ・ Loto7当選番号予想
- ・ サッカー選手のゴール数の予測



# Loto7 当選番号予想



# 背景と方法

## ▶ 背景

宝くじは競馬やパチンコ、株などに比べて実力の影響がないため、回収率が低いという情報を手に入れた。

本当に回収率が低いのか、どうにか実力の影響を出す方法を発見し、当選することはできないかと思ったため。

## ▶ 方法

第1回から400回までの当選番号を抽出して、それぞれの数字の出る確率を求める。

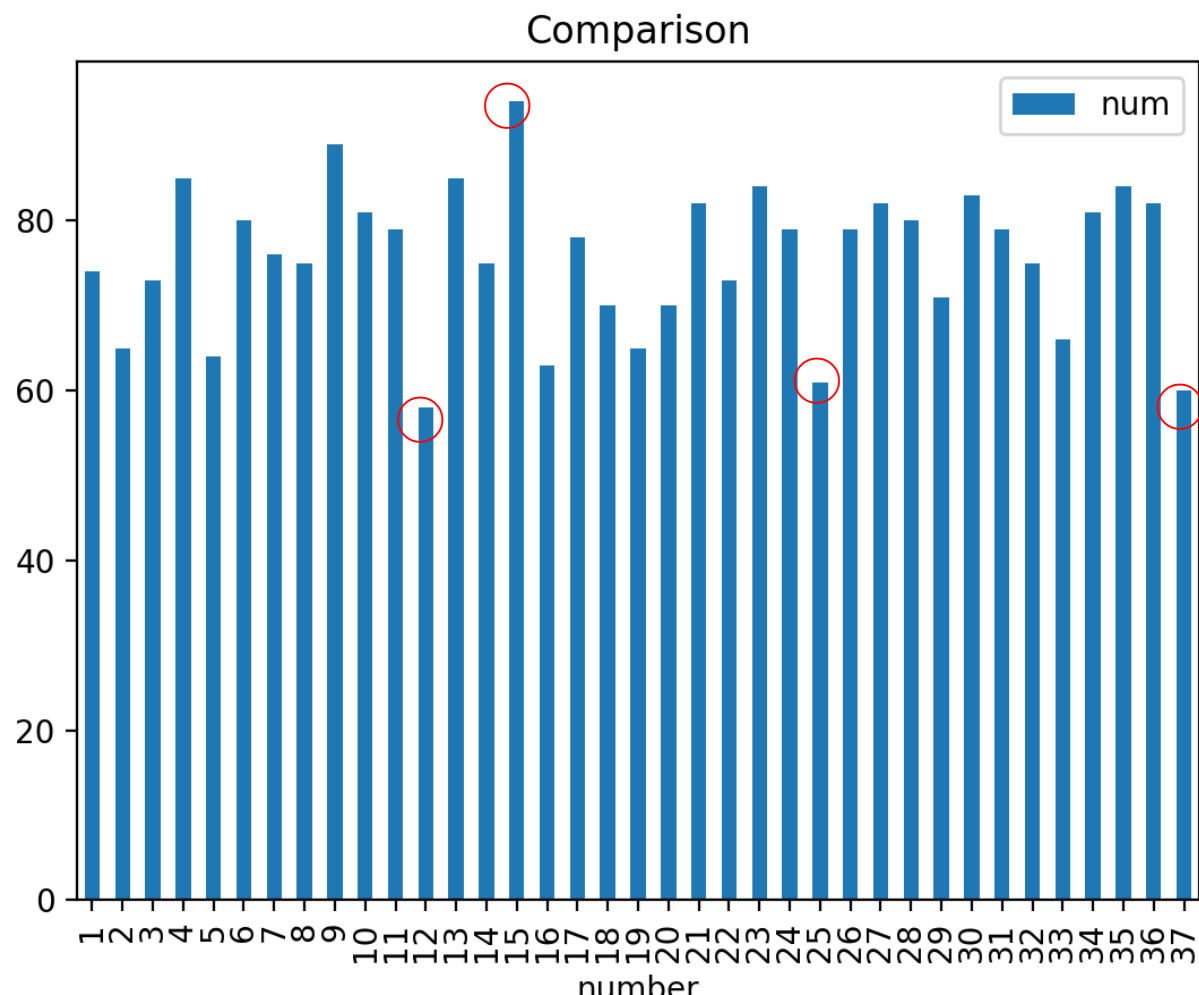
これらの確率通りに7つの数字を出力するプログラムを作る。

# データ収集

- Seleniumでデータ収集  
(chromedriverを用いる)
- 本数字のclassの要素  
を取り出し、それぞれ  
の数字のtagのテキス  
トを取得する

本数字	10	13	17	18	22	28	37
-----	----	----	----	----	----	----	----

# それぞれの数字の出た回数



- ▶ 結果は右図のようになる。
- ▶ 最大値をとるのは1 5で、最小値をとるのは1 2、2 5、3 7あたりである
- ▶ 1つの数字が出る回数  $7 \times 400 / 37 = 76$ 回



# 当選番号の予想

- ・ それぞれの数字が出る確率を求めて、求めた確率通りにランダムで7つの数字を出力するプログラムを作る

# 結果

## 当選 0/5回中

- ・自分の考えたやり方では実力の影響が出なかった。
- ・実力の影響がでなかったため、回収率が低いという情報に賛同できる。



# サッカー選手 のゴール数の 予測

# 背景と方法

## ▶ 背景

あるサッカー選手はドリブルからのシュートの決定力があると友達と話したことから、ドリブルなどのデータから点を決めるか予測できるのかと思い、機械学習を用いてゴールの予測するプログラムを作る。

## ▶ 方法

WhoScoredというサイトからリオネル・メッシ選手のオフエンス面でのデータからゴール数とゴールするかを予測する。

※Whoscoredのurl

<https://www.whoscored.com/Players/11119/MatchStatistics/Lionel-Messi>

# データの収集

- ・ Seleniumでデータ収集(chromedriverを用いる)

- ・ 2020年9月2日から2021年5月2日までのデータを収集する。

## データ情報

Mins: プレー時間

Goals: 総ゴール数

Assists: 総アシスト数

Shots: 総シュート数

KeyP: キーパスの総数

Drb: 成功したドリブル

Fouled: ファールを受けた回数

Off: オフサイドをとられた

Disp: プレイヤーが相手にボールを奪われた-ドリブルはしていない

UnsTch: ボールコントロールにおけるミスでボールを失った回数。

# 結果予測

## ▶ ゴール数を予測

2021年5月3日以降のデータを用いて  
ゴール数を予測する  
(Randomforestを採用)

## ▶ ゴールを決めたか決めてないかを予測

2021年5月3日以降の  
データを用いてゴールを  
決めるか決めないかを予  
測する。  
(LGBMClassifierを採用)

## 結果「ゴール数の予測」

	結果	予測
2021/5/8	0	0
2021/5/12	1	0
2021/5/17	1	1
2021/5/23	0	0

正解率  
75%

※5/23日は試合に出てないので出場時間(Min)を90、それ以外のデータを0にした

# 結果「ゴールを決めたか決めてないかを予測」

	結果	予測	
2021/5/8	0	1	
2021/5/12	1	1	
2021/5/17	1	1	正解率
2021/5/23	0	0	75%

※5/23日は試合に出てないので出場時間(Min)を90、それ以外のデータを0にした

※決めた場合 1  
決めなかった場合 0



## 結果から

- ・ゴール数の予測が0に対し、ゴールを決めたかの予測で1と矛盾が応じるところがあった。これは、ゴール数の予測した値を小数点を切り捨てしているのに対して、ゴールを決めたかを1か0かで分類する際に、少数の値が四捨五入(例  $0.51 \rightarrow 1, 0.50 \rightarrow 0$ )されるためであると推測。
- ・結果と予測が合わなかったところは、たくさんの決定機はあったがゴールを決めることができないことがあったためであると推測。

# 今後の予定

- ・ 来シーズンも引き続き予測を行う。
- ・ どのデータがゴール数と関係性が高いかを調べる。
- ・ 予測した値の少数点の処理に違うやり方を採用してみる。  
(例  $x$ が50%以上  $\rightarrow 1$ ,  $\rightarrow$   
 $x$ が50%未満  $\rightarrow 0$ )  
 $x$ が80%以上  $\rightarrow 1$ ,  
 $x$ が80%未満  $\rightarrow 0$
- ・ Kaggleで実力をつける

# codeや詳しい内容

- github アカウント  
tasu0630

※主に製作物やkaggleなどのプログラムを記録

- Kaggle アカウント  
tasu0630

※参加した企画の順位が記載

以上が活動実績となります  
ありがとうございました

