

# 夏合宿課題

今年のプロ野球  
優勝チームはどこだ



ニューラルネットを用いて  
セリーグの順位予想を試してみた

初めにあまり野球を見ない人もいると思うので

プロ野球には12球団があり、2つのリーグが存在する



セ・リーグ



パ・リーグ

今回予想する優勝チームはこっち

# 12球団もあるのでそれぞれのチームに特色がある

- ①得点力はあるが投手力がない
  - ②投手力があるが得点力がない
  - ③両方まずまず
  - ④長打力はないがアベレージは高い
  - ⑤監督の采配が冴えている
  - ⑥エラー、四死球が多い
- などなど、、、、、、

# 予測の仕方

- ・ニューラルネットを用いて、セリーグ6球団の勝利数を出力値として出す
- ・使用する説明変数
  - 去年のチーム打率
  - 去年のチーム防御率
  - オープン戦の順位
  - 去年のチームにホームラン王がいたか

# 使用する説明変数の用語説明

- 首位打者（リーグで一番打率が高い選手のこと。簡単に言えば毎回の打席でヒットを打つ確率が一番高い人。）
- オープン戦（公式戦の前に行う練習試合的なもの、順位とかもはっきり出る。）
- 交流戦（6月後半から7月半ごろに行う、セリーグ対パリーグで行う試合）
- 最優秀防御率（簡単に言ったら点を取られない先発ピッチャー）
- ホームラン王（ホームランを一番うった選手）

教科書では牛乳の販売個数を予測値として出す式があってそれを使えばうまくいくと思ったが全くうまくいかず、、、

The screenshot shows the RStudio interface with the following R code in the editor:

```
16 open1 = (yakult[1:20,2]-min(yakult[1:20,2]))/(max(yakult[1:20,2])-min(yakult[1:20,2]))
17 title1 = (yakult[1:20,3]-min(yakult[1:20,3]))/(max(yakult[1:20,3])-min(yakult[1:20,3]))
18 bougyo1 = (yakult[1:20,4]-min(yakult[1:20,4]))/(max(yakult[1:20,4])-min(yakult[1:20,4]))
19 daritu1 = (yakult[1:20,5]-min(yakult[1:20,5]))/(max(yakult[1:20,5])-min(yakult[1:20,5]))
20 inp2 = cbind(open1,title1,bougyo1,daritu1)
21
22 tekiyou = predict(study,inp2)
23 tekiyou
24
25 kekka =round(tekiyou)
26
27 kekka1 = kekka*(max(yakult[1:19,6])-min(yakult[1:19,6]))+min(yakult[1:19,6])
28 kekka1
29
30
```

The console output shows the following values for the predicted variable:

```
R 4.2.3 ~./Desktop/
[3,] 87
[4,] 87
[5,] 87
[6,] 87
[7,] 87
[8,] 87
[9,] 55
[10,] 87
[11,] 87
[12,] 55
[13,] 55
[14,] 87
```

The Environment pane on the right shows the following objects:

Object	Class	Attributes
inp2	num	[1:20, 1:4] 0.182 0.273 0...
kekka	num	[1:20, 1] 1 1 1 1 1 1 1...
kekka1	num	[1:20, 1] 87 87 87 87 87 ...
study	List of 15	
tekiyou	num	[1:20, 1] 0.574 0.574 0.5...
yakult	20 obs. of 6 variables	

The bottom part of the image shows a presentation slide with a classification tree diagram:

(結果)目的変数	
量的変数	質的変数
数量化I類	数量化II類
性別と菓物の好き嫌いから	性別と菓物の好き嫌いから

やり方わかる人いたらぜひ教えてください！



気を取り直して、  
プロ野球データにおけるいろいろな連関規則  
について調べてみた

- 優勝するチームにはどんな特徴があるのか？  
→ここから今年優勝しそうなチームを予測（今の所多分阪神だけ  
ど、、今が4月だと仮定して）
- 最下位になるチームはどんなチームか？
- 昨年ホームラン王がいたチームは今年どうなるのか？  
……などなど

# 使用したcsvファイル

自分が生まれた2003年から2022年まで、下のような説明変数でtrueなら、falseなら0で記録。(優勝する→1(true))

baseball												
セリーグチーム	優勝する	最下位	昨年1位	昨年Aクラス	昨年最下位	監督が昨年から変わった	昨年首位打者いる	昨年ホームラン王	昨年最優秀防御率	オープン戦で3位以上	オープン戦で10位以下	交流戦で5位以上
2003(阪神)	1	0	0	0	1	0	0	0	0	1	0	0
2003(ヤクルト)	0	0	0	1	0	0	0	0	0	0	1	0
2003(横浜)	0	1	0	0	1	1	0	0	0	0	0	0
2003(広島)	0	0	0	0	0	0	0	0	0	1	0	0
2003(巨人)	0	0	1	1	0	0	0	1	1	0	0	0
2003(中日)	0	0	0	1	0	1	1	0	0	1	0	0
2004(ヤクルト)	0	0	0	0	0	0	0	1	0	0	0	0
2004(阪神)	0	0	1	1	0	0	1	0	1	1	0	0
2004(横浜)	0	1	0	0	1	0	0	1	0	0	1	0
2004(広島)	0	0	0	0	0	0	0	0	0	0	0	0
2004(巨人)	0	0	0	1	0	1	0	0	0	0	0	0
2004(中日)	1	0	0	1	0	0	0	0	0	0	0	0
2005(広島)	0	1	0	0	0	0	1	0	0	0	1	0
2005(阪神)	1	0	0	0	0	0	0	0	0	0	0	1
2005(横浜)	0	0	0	0	1	1	0	1	0	1	0	0
2005(巨人)	0	0	0	1	0	0	0	1	1	0	0	1
2005(中日)	0	0	1	1	0	0	0	0	0	0	0	0
2005(ヤクルト)	0	0	0	1	0	0	0	0	0	0	1	1
2006(中日)	1	0	0	1	0	0	0	0	0	1	0	1
2006(阪神)	0	0	1	1	0	0	0	0	0	1	0	1
2006(横浜)	0	1	0	1	0	0	0	0	1	0	0	0
2006(巨人)	0	0	0	0	0	0	0	0	0	0	0	0
2006(ヤクルト)	0	0	0	0	0	1	1	0	0	0	1	1
2006(広島)	0	0	0	0	1	1	0	1	0	0	0	0
2007(ヤクルト)	0	1	0	1	0	0	0	0	0	1	0	0
2007(阪神)	0	0	0	1	0	0	0	0	0	0	0	0
2007(横浜)	0	0	0	0	1	1	0	0	0	0	1	1
2007(広島)	0	0	0	0	0	0	0	0	1	0	0	0
2007(巨人)	1	0	0	1	0	0	0	0	0	0	0	1
2007(中日)	0	0	1	0	0	0	1	1	0	0	0	1
2008(広島)	0	0	0	0	0	0	0	0	0	0	0	0
2008(巨人)	1	0	1	1	0	0	0	0	1	0	1	1
2008(ヤクルト)	0	0	0	0	1	1	1	0	0	0	0	0
2008(阪神)	0	0	0	1	0	0	0	0	0	0	0	1
2008(中日)	0	0	0	1	0	0	0	0	0	1	0	0
2008(横浜)	0	1	0	0	0	0	0	1	0	0	1	0
2009(巨人)	1	0	1	1	0	0	0	0	0	0	0	0
2009(ヤクルト)	0	0	0	0	0	0	0	0	1	1	0	1
2009(阪神)	0	0	0	1	0	1	0	0	0	0	1	0
2009(中日)	0	0	0	1	0	0	0	0	0	0	1	0
2009(横浜)	0	1	0	0	1	1	1	1	0	0	0	0
2009(広島)	0	0	0	0	0	0	0	0	0	0	0	1
2010(中日)	1	0	0	1	0	0	0	1	1	0	0	0
2010(巨人)	0	0	1	1	0	0	1	0	0	0	0	0

☐ — ト

```
yusyou = read.csv("baseball.csv",header=TRUE,row.names=1)
yusyou = as.matrix(yusyou)
yusyoutr = as(yusyou,"transactions")
summary(yusyoutr)
inspect(yusyoutr)
itemFrequency(yusyoutr)
itemFrequencyPlot(yusyoutr,ylim =c(0,1))
yusyourule = apriori(yusyoutr,parameter =
list(maxlen=4,support=0.04,confidence=0.3,ext=TRUE))
inspect(yusyourule)
yusyoudairule =subset(yusyourule,subset=(rhs %in% "???"&(lift>1.0)))
inspect(yusyoudairule)
```

# まず、優勝するチームの特徴は？

lhs	rhs	support	confidence	coverage	lift	count
[1] {昨年1位, オープン戦で10位以下}	=> {優勝する}	0.04166667	0.8333333	0.05000000	5.0	5
[2] {昨年1位, 交流戦で5位以上}	=> {優勝する}	0.05000000	0.6666667	0.07500000	4.0	6
[3] {昨年Aクラス, 交流戦で5位以上}	=> {優勝する}	0.07500000	0.5000000	0.15000000	3.0	9

セ・リーグ CENTRAL LEAGUE						
順位	チーム	試	勝	敗	分	差
1	ヤクルト	143	80	59	4	-
2	DeNA	143	73	68	2	8
3	阪神	143	68	71	4	4
4	巨人	143	68	72	3	0.5
5	広島	143	66	74	3	2
6	中日	143	66	75	2	0.5

図1.2022年セ・リーグ順位

DeNAワンチャンありそう???

順位	チーム名	試合	勝利	敗戦	引分	勝率	勝差	試合	得点	失点	塁打	塁塁	打率	御率	大策
1	DeNA	18	11	7	0	.611	-	0	80	59	8	5	.258	2.93	9
2	ソフトバンク	18	11	7	0	.611	0	0	86	65	16	6	.266	3.30	8
3	巨人	18	11	7	0	.611	0	0	72	55	25	9	.274	3.00	5
4	オリックス	18	11	7	0	.611	0	0	66	56	15	7	.232	2.75	8
5	日本ハム	18	10	8	0	.556	1	0	59	45	19	10	.238	2.24	10
6	楽天	18	9	9	0	.500	1	0	65	89	14	8	.260	4.29	12
7	広島	18	9	9	0	.500	0	0	61	76	7	13	.238	3.87	13
8	ロッテ	18	7	9	2	.438	1	0	55	77	13	13	.215	4.08	8
9	中日	18	7	10	1	.412	0.5	0	54	58	10	11	.232	2.81	11
10	阪神	18	7	10	1	.412	0	0	58	66	5	13	.210	3.23	17
11	ヤクルト	18	7	11	0	.389	0.5	0	69	65	13	5	.251	3.53	6
12	西武	18	6	12	0	.333	1	0	48	62	11	8	.212	3.17	10

図2.今年の交流戦順位

# 今年のAクラスを予測

まず先ほどのスライドでDeNAのAクラスかつ優勝を予測

lhs	rhs	support	confidence	coverage	lift	count
[1] {昨年最優秀防御率}	=> {Aクラス}	0.10000000	0.6000000	0.16666667	1.200000	12
[2] {オープン戦で3位以上}	=> {Aクラス}	0.13333333	0.6956522	0.19166667	1.391304	16
[4] {昨年ホームラン王}	=> {Aクラス}	0.10000000	0.5217391	0.19166667	1.043478	12
[5] {昨年1位}	=> {Aクラス}	0.14166667	0.8500000	0.16666667	1.700000	17
[6] {交流戦で5位以上}	=> {Aクラス}	0.18333333	0.6666667	0.27500000	1.333333	22
[7] {昨年Aクラス}	=> {Aクラス}	0.30833333	0.6166667	0.50000000	1.233333	37
<del>[8] {昨年Aクラス, オープン戦で3位以上}</del>	<del>=&gt; {Aクラス}</del>	<del>0.08333333</del>	<del>0.6666667</del>	<del>0.12500000</del>	<del>1.333333</del>	<del>10</del>
<del>[12] {昨年1位, 交流戦で5位以上}</del>	<del>=&gt; {Aクラス}</del>	<del>0.07500000</del>	<del>1.0000000</del>	<del>0.07500000</del>	<del>2.000000</del>	<del>9</del>
<del>[14] {昨年Aクラス, 交流戦で5位以上}</del>	<del>=&gt; {Aクラス}</del>	<del>0.11666667</del>	<del>0.7777778</del>	<del>0.15000000</del>	<del>1.555556</del>	<del>14</del>

この中でsupportとconfidenceの両方が高いものを選ぶと、、、、

この二つにあてはまるのは阪神、ヤクルト

ちなみに[1]に当てはまるのは阪神、[2]にあてはまるのは巨人、[4]に当てはまるのはヤクルト



# 現在の順位

DeNAが優勝するのはだいぶ厳しいです

順位	チーム名	試合	勝利	敗戦	引分	勝率	勝差	残試合	得点	失点	本塁打	盗塁	打率	防御率	失策
1	 阪神	121	73	44	4	.624	<b>M14</b>	22	480	368	67	66	.247	2.72	71
2	 広島	124	67	53	4	.558	7.5	19	439	434	84	69	.246	3.13	65
3	 DeNA	121	61	57	3	.517	5	22	461	428	91	26	.251	3.24	57
4	 巨人	121	60	60	1	.500	2	22	469	466	145	41	.258	3.69	46
5	 ヤクルト	122	48	71	3	.403	11.5	21	448	492	101	56	.238	3.72	65
6	 中日	121	46	72	3	.390	1.5	22	338	432	57	32	.241	3.19	71

ここからは番外編です



# 成績が悪く監督やめるとよく言われるが監督が変わった次の年はどんな成績になりやすいのか

lhs	rhs	support	confidence	coverage	lift	count
[1] {監督が昨年から変わった}	=> {Bクラス}	0.14166667	0.56666667	0.25000000	1.1333333	17
[2] {監督が昨年から変わった, 昨年首位打者いる}	=> {Bクラス}	0.04166667	0.6250000	0.06666667	1.250000	5
[3] {昨年最下位, 監督が昨年から変わった}	=> {Bクラス}	0.08333333	0.6666667	0.12500000	1.3333333	10
[4] {監督が昨年から変わった, オープン戦で10位以下}	=> {Aクラス}	0.04166667	0.5555556	0.07500000	1.1111111	5
[5] {昨年Aクラス, 監督が昨年から変わった}	=> {Aクラス}	0.04166667	0.7142857	0.05833333	1.428571	5

成績が悪いからと言って監督が変わっても次の年はあまり期待できない！  
確かに翌々年にチームが再建する例はよく見るが、翌年に再建する例は稀

# 個人の成績がいい選手がいるチームはど うなる？

昨年のホームラン王がいるチーム

lhs	rhs	support	confidence	coverage	lift	count
[1] {昨年ホームラン王}	=> {Aクラス}	0.10000000	0.5217391	0.19166667	1.043478	12
[2] {昨年Aクラス, 昨年ホームラン王}	=> {Aクラス}	0.05833333	0.5833333	0.10000000	1.166667	7

昨年の首位打者がいるチーム

lhs	rhs	support	confidence	coverage	lift	count
[1] {昨年首位打者いる}	=> {Bクラス}	0.10833333	0.6500000	0.16666667	1.300000	13
[2] {監督が昨年から変わった, 昨年首位打者いる}	=> {Bクラス}	0.04166667	0.6250000	0.06666667	1.250000	5
[3] {昨年Aクラス, 昨年首位打者いる}	=> {Bクラス}	0.04166667	0.5555556	0.07500000	1.111111	5

昨年の最優秀防御率がいるチーム

lhs	rhs	support	confidence	coverage	lift	count
[1] {昨年最優秀防御率}	=> {Aクラス}	0.10000000	0.6000000	0.16666667	1.200000	12
[2] {昨年最優秀防御率, 交流戦で5位以上}	=> {Aクラス}	0.04166667	0.7142857	0.05833333	1.428571	5
[3] {昨年Aクラス, 昨年最優秀防御率}	=> {Aクラス}	0.05833333	0.5833333	0.10000000	1.166667	7

# まとめ

- プロ野球はデータによる予測が当たることもあれば、全く正反對の結果になったりもするのでそこが面白いと思った。
- 現代では、スポーツやギャンブルなどの予測をデータを見て楽しむことも増えてきているので今回それを見にしみて体感した。