資料探勘HW3

AI三 B1228005 胡樂麒

1.輸出結果

下表為本次實際跑出的準確率（Accuracy），對應程式順序：

編號	特徵組合/模型	準確率
(A)	[HomeLastWin, VisitorLastWin]/ 兩隊是否在「上一場比賽」贏球/ 決策樹分類	58.0%
(B)	["HomeLastWin","VisitorLastWin","HomeTeamRanksHigher"] 加入「主隊在 2013 年排名是否較高」。 / 決策樹分類	60.3%
(C)	[HomeTeamRanksHigher, HomeTeamWonLast]加入「兩隊上次交手主隊是否勝出」。 / 決策樹分類	60.6%
(D)	One-Hot 編碼後的球隊名稱（X_teams_expanded）將球隊名稱轉成 one-hot 向量，讓決策樹能從「特定球隊」中學習。/ 決策樹分類	60.7%
(E)	Label Encoded 球隊（X_teams）隨機森林取代單棵決策樹，特徵為「整數型球隊代號」。/ 隨機森林分類	59.2%
(F)	[HomeLastWin, VisitorLastWin, HomeTeamRanksHigher] + X_teams(Label)合併戰績特徵與球隊編碼，理論上資訊更多，但成效變弱。 / 隨機森林分類	58.5%
(G)	RandomForestClassifier + GridSearchCV 使用網格搜尋找到最佳參數組合（entropy、max_features=10、min_samples_leaf=6），提升穩定度。/ 隨機森林分類	60.6%
(H)	新增進階特徵（休息天數、近五場、客隊在主場）新增「休息天數、近五場勝率、客隊在主場表現」。但準確率降低。/ 隨機森林分類	57.3%

輸出結果截圖

```
PS C:\Users\User\Downloads\資料探勘hw3> python -u "c:\Users\User\Downloads\資料探勘hw3\B1228005 HW3.py"
         Date          Visitor Team  VisitorPts           Home Team  HomePts  OT? Score Type Notes
0  2013-10-29        Orlando Magic          87      Indiana Pacers       97  NaN  Box Score   NaN
1  2013-10-29        Chicago Bulls          95         Miami Heat      107  NaN  Box Score   NaN
2  2013-10-29  Los Angeles Clippers         103   Los Angeles Lakers      116  NaN  Box Score   NaN
3  2013-10-30        Brooklyn Nets          94  Cleveland Cavaliers       98  NaN  Box Score   NaN
4  2013-10-30       Boston Celtics          87     Toronto Raptors       93  NaN  Box Score   NaN
         Date          Visitor Team  VisitorPts           Home Team  HomePts  OT? Score Type Notes
0 2013-10-29        Orlando Magic          87      Indiana Pacers       97  NaN  Box Score   NaN
1 2013-10-29        Chicago Bulls          95         Miami Heat      107  NaN  Box Score   NaN
2 2013-10-29  Los Angeles Clippers         103   Los Angeles Lakers      116  NaN  Box Score   NaN
3 2013-10-30        Brooklyn Nets          94  Cleveland Cavaliers       98  NaN  Box Score   NaN
4 2013-10-30       Boston Celtics          87     Toronto Raptors       93  NaN  Box Score   NaN
         Date          Visitor Team  VisitorPts           Home Team  HomePts  OT? Score Type Notes  HomeWin HomeLastWin VisitorLastWin
20 2013-11-01  New Orleans Pelicans          90       Orlando Magic      110  NaN  Box Score   NaN     True       False          False
21 2013-11-01    Philadelphia 76ers         109  Washington Wizards      102  NaN  Box Score   NaN    False       False           True
22 2013-11-01      Toronto Raptors          95       Atlanta Hawks      102  NaN  Box Score   NaN     True       False           True
23 2013-11-01      Milwaukee Bucks         105      Boston Celtics       98  NaN  Box Score   NaN    False       False          False
24 2013-11-01           Miami Heat         100       Brooklyn Nets      101  NaN  Box Score   NaN     True       False          False
Accuracy: 58.0%
```

	Rk	Team	Overall	Home	Road	E	W	A	C	...	≤3	≥10	Oct	Nov	Dec	Jan	Feb	Mar	Apr
0	1	Miami Heat	66-16	37-4	29-12	41-11	25-5	14-4	12-6	...	9-3	39-8	1-0	10-3	10-5	8-5	12-1	17-1	8-1
1	2	Oklahoma City Thunder	60-22	34-7	26-15	21-9	39-13	7-3	8-2	...	3-6	44-6	NaN	13-4	11-2	11-5	7-4	12-5	6-2
2	3	San Antonio Spurs	58-24	35-6	23-18	25-5	33-19	8-2	9-1	...	9-5	31-10	1-0	12-4	12-4	12-3	8-3	10-4	3-6
3	4	Denver Nuggets	57-25	38-3	19-22	19-11	38-14	5-5	10-0	...	11-7	28-8	0-1	8-8	9-6	12-3	8-4	13-2	7-1
4	5	Los Angeles Clippers	56-26	32-9	24-17	21-9	35-17	7-3	8-2	...	3-5	38-12	1-0	8-6	16-0	9-7	8-5	7-7	7-1
5	6	Memphis Grizzlies	56-26	32-9	24-17	22-8	34-18	8-2	8-2	...	6-4	28-9	0-1	12-1	7-7	10-7	9-2	11-6	7-2
6	7	New York Knicks	54-28	31-10	23-18	37-15	17-13	10-6	12-6	...	7-5	31-12	NaN	11-4	10-5	7-6	6-5	12-6	8-2
7	8	Brooklyn Nets	49-33	26-15	23-18	36-16	13-17	11-5	13-5	...	9-4	23-17	NaN	11-4	5-11	11-4	7-5	8-7	7-2
8	9	Indiana Pacers	49-32	30-11	19-21	31-20	18-12	6-11	13-3	...	4-9	27-14	1-0	7-8	10-5	9-6	9-3	11-5	2-5
9	10	Golden State Warriors	47-35	28-13	19-22	19-11	28-24	7-3	5-5	...	5-3	20-18	1-0	8-6	12-4	8-7	4-8	9-7	5-3
10	11	Chicago Bulls	45-37	24-17	21-20	34-18	11-19	13-5	9-7	...	11-7	16-16	1-0	6-7	9-6	12-4	5-8	7-7	5-5
11	12	Houston Rockets	45-37	29-12	16-25	21-9	24-28	7-3	7-3	...	5-5	26-13	1-0	6-8	10-6	8-9	6-5	9-5	5-4
12	13	Los Angeles Lakers	45-37	29-12	16-25	17-13	28-24	6-4	6-4	...	8-5	18-17	0-2	8-6	7-7	5-11	9-4	9-6	7-1
13	14	Atlanta Hawks	44-38	25-16	19-22	29-23	15-15	7-11	11-7	...	5-5	19-20	NaN	9-5	10-5	7-9	7-4	8-10	3-5
14	15	Utah Jazz	43-39	30-11	13-28	17-13	26-26	5-5	5-5	...	5-7	19-21	1-0	8-8	6-9	10-4	6-6	7-9	5-3
15	16	Boston Celtics	41-40	27-13	14-27	27-24	14-16	7-9	8-9	...	8-7	18-23	0-1	9-6	5-9	8-7	8-4	8-8	3-5
16	17	Dallas Mavericks	41-41	24-17	17-24	17-13	24-28	5-5	6-4	...	5-8	17-19	1-1	6-8	5-10	7-8	6-5	11-5	5-4
17	18	Milwaukee Bucks	38-44	21-20	17-24	24-28	14-16	11-7	7-9	...	7-5	13-25	NaN	7-7	9-6	8-7	4-8	7-9	3-7
18	19	Philadelphia 76ers	34-48	23-18	11-30	22-30	12-18	7-9	7-11	...	4-5	13-24	1-0	9-6	4-11	5-9	3-8	8-9	4-5
19	20	Toronto Raptors	34-48	21-20	13-28	22-30	12-18	5-11	8-10	...	8-8	16-22	0-1	4-12	7-7	5-10	7-5	4-11	7-2
20	21	Portland Trail Blazers	33-49	22-19	11-30	15-15	18-34	5-5	5-5	...	9-6	13-24	1-0	5-10	9-4	8-8	3-9	7-9	0-9
21	22	Minnesota Timberwolves	31-51	20-21	11-30	14-16	17-35	4-6	7-3	...	3-10	15-26	NaN	7-8	7-5	3-12	3-10	6-11	5-5
22	23	Detroit Pistons	29-53	18-23	11-30	25-27	4-26	6-12	8-8	...	6-10	15-29	0-1	5-11	6-10	6-7	6-8	1-13	5-3
23	24	Washington Wizards	29-53	22-19	7-34	15-37	14-16	5-13	5-13	...	6-9	13-17	0-1	1-12	3-11	7-9	7-5	9-8	2-7
24	25	Sacramento Kings	28-54	20-21	8-33	14-16	14-38	4-6	4-6	...	7-3	12-31	0-1	4-10	7-8	6-11	3-9	7-8	1-7
25	26	New Orleans Hornets	27-55	16-25	11-30	12-18	15-37	3-7	5-5	...	7-5	10-27	0-1	4-9	3-13	8-8	5-8	6-9	1-7
26	27	Phoenix Suns	25-57	17-24	8-33	8-22	17-35	1-9	4-6	...	6-8	10-31	0-1	7-9	4-11	5-9	4-9	3-12	2-6
27	28	Cleveland Cavaliers	24-58	14-27	10-31	18-34	6-24	5-13	3-13	...	6-11	7-27	1-0	3-12	3-13	6-8	7-5	2-12	2-8
28	29	Charlotte Bobcats	21-61	15-26	6-35	18-34	3-27	6-12	6-12	...	6-6	6-37	NaN	7-8	1-15	3-11	2-10	4-12	4-5
29	30	Orlando Magic	20-62	12-29	8-33	10-42	10-20	2-16	5-13	...	2-9	8-30	NaN	5-10	7-9	2-12	2-11	3-13	1-7

```
[30 rows x 24 columns]
Accuracy: 60.3%
Accuracy: 60.6%
Accuracy: 60.7%
Accuracy: 59.2%
Accuracy: 58.5%
Accuracy: 60.6%
RandomForestClassifier(criterion='entropy', max_features=10, min_samples_leaf=6,
                       random_state=14)

--- Adding New Engineered Features ---

Accuracy (with new engineered features): 57.3%
PS C:\Users\User\Downloads\資料探勘hw3>
```

2.自己新增特徵（Engineered Features）

1. 休息天數：DaysSinceHome、DaysSinceVisitor（本場日期 − 該隊上一次出賽日的天數）。
2. 近五場勝率：HomeRecentWinRatio、VisitorRecentWinRatio（各隊過去 5 場勝場比例，若無歷史則以 0.5）。
3. 客隊在特定主場的歷史勝率：VisitorVenueWinRate（key =（visitor, home）的歷史對戰中，客隊勝率；若無歷史則 0.5）。

3. 結論

本作業以決策樹（Decision Tree Classifier） 與 隨機森林（Random Forest Classifier） 兩種監督式學習演算法，對 NBA 2013–2014 年度比賽進行主場勝負預測。透過多階段特徵工程與交叉驗證的比較實驗，結果顯示：

1. 基礎特徵（HomeLastWin、VisitorLastWin） 約 58% 的預測準確率，顯示單一比賽結果的延續性有限。

2. 加入 排名高低（HomeTeamRanksHigher） 與 上次交手結果（HomeTeamWonLast） 後，準確率提升至約 60%–61%，說明球隊整體實力與歷史對戰關係對結果具有顯著影響。

3. 將球隊以 One-Hot 編碼 形式輸入模型時，決策樹能進一步捕捉球隊特性，準確率達約 60.7%。

4. 使用 隨機森林 進行集成學習可有效穩定預測結果，而透過 Grid Search 調整參數後（criterion='entropy', max_features=10, min_samples_leaf=6），模型準確率達到 60.6% 。

5. 新增的進階特徵（休息天數、近五場勝率、客隊在特定主場勝率）雖提升了模型的資訊量，但其表現未優於基準模型（約 57.3%）。然而，這部分仍展現了特徵工程對模型改進的重要性。

在本次 NBA 主場勝負預測的作業中，從準確率可以看出，球隊的「整體實力（排名）」與「歷史對戰關係」是提升模型準確率的關鍵。表現最好的模型在準確率上達到了 60.7%。儘管進階特徵並未成功優化模型，但也顯現了特徵工程對預測結果有著至關重要影響。