Predicting IAAF World Half Marathon Championship **Training Strategy for** Athlete

Ruofan Lyu 001222480

### Introduction

IAAF(International Association of Athletics Federations) World Half Marathon Championship is a well-known sports event in the world. As a possible pre-qualification of the most famous six World Marathon Major, including Boston Marathon, it is a highly participated marathon event.

Predicting athletes' performance is significant. It helps athletes to determine their further training strategy. For sport enthusiasts, predicting can bring more professional guidance to them and attract more public attention.

# Approach

- Web data scraping
- Build a regression model based on predictor variables: age, discipline, performance, date

```
=> IAAF_code | Name | Nationality | discipline | performance | date | 273114 | MOHAMUD AADAN | GREAT BRITAIN & NI | 3000 M | 8:13.12 | 7/15/2017
```

Provide a training strategy

# Multiple regression

In the multivariate case, when there is more than one independent variable, the regression line cannot be visualized in the two dimensional space. We could construct a linear equation containing all those variables. In general then, multiple regression procedures will estimate a linear equation of the form:

$$Y = a + b_1 * X_1 + b_2 * X_2 + ... + b_p * X_p$$

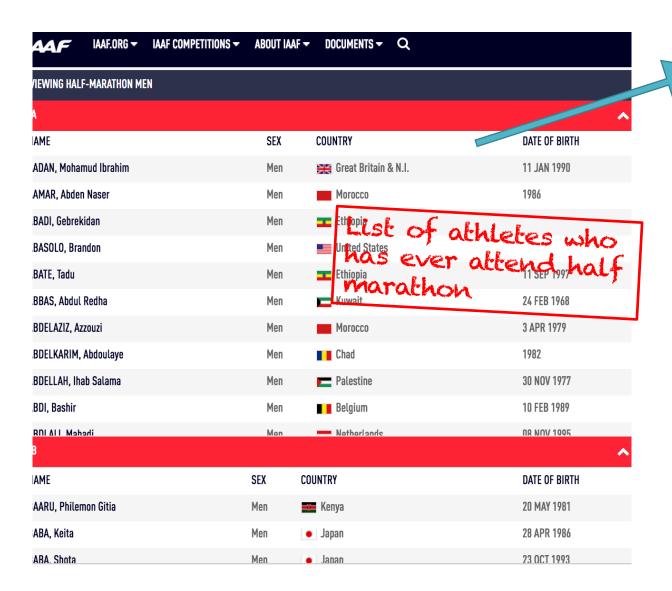
# Naive Bayes classifier

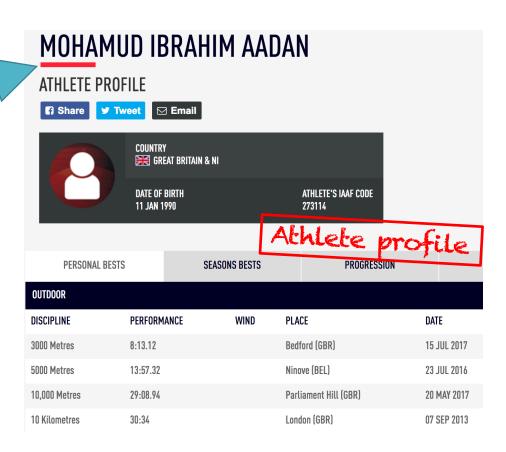
In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with independence assumptions between the features.

$$P(C_k|x)\frac{P(C_k)P(x|C_k)}{P(x)}$$

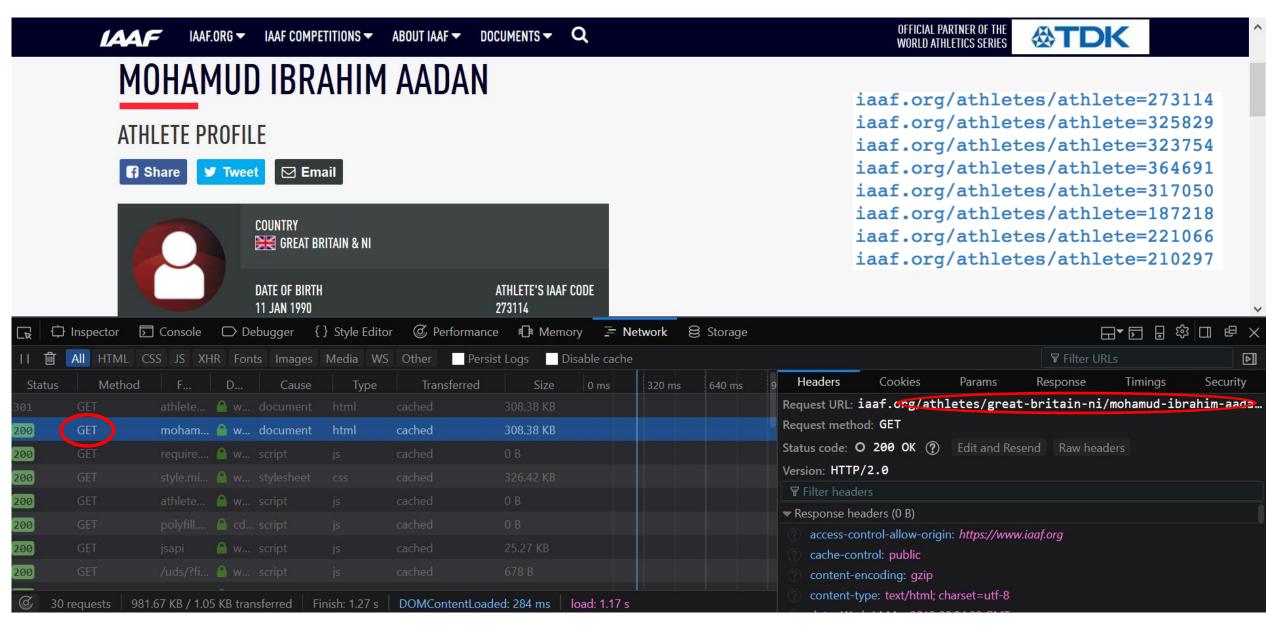
Knowing the probability of  $x \mid Ck$ , calculate  $P(Ck \mid x)$ : to determine the result.

## Data Source





### Data scraping



### Redirected URL $\rightarrow$

#### Original URL



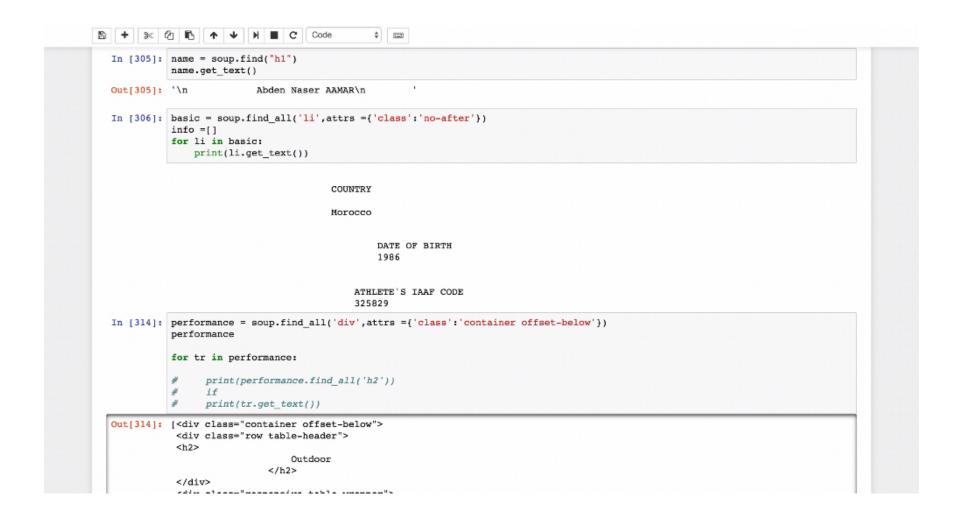
#### t(href)

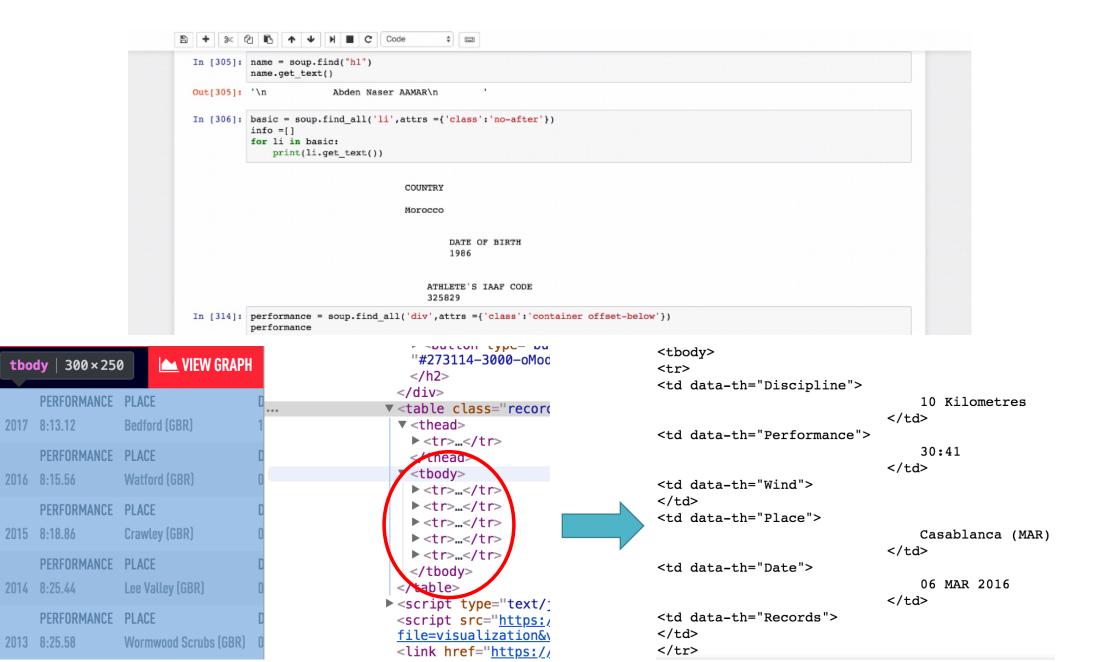
```
iaaf.org/athletes/athlete=273114
iaaf.org/athletes/athlete=325829
iaaf.org/athletes/athlete=323754
iaaf.org/athletes/athlete=364691
iaaf.org/athletes/athlete=317050
iaaf.org/athletes/athlete=187218
iaaf.org/athletes/athlete=221066
iaaf.org/athletes/athlete=210297
iaaf.org/athletes/athlete=56172
iaaf.org/athletes/athlete=281105
iaaf.org/athletes/athlete=350387
iaaf.org/athletes/athlete=263903
iaaf.org/athletes/athlete=137596
iaaf.org/athletes/athlete=242055
iaaf.org/athletes/athlete=243638
iaaf.org/athletes/athlete=136290
iaaf.org/athletes/athlete=251403
iaaf.org/athletes/athlete=138182
iaaf.org/athletes/athlete=310866
inof one/othlotos/othloto-2000E0
```

```
# identify user agent to get access to the GET request
headers = {'User-Agent':'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_12_6) Apr
for i in urls:
    rs = requests.get(i,headers = headers).url
    links.append(rs)
    print(rs)

https://www.iaaf.org/athletes/great-britain-ni/mohamud-ibrahim-aadan-273114
```

```
https://www.iaaf.org/athletes/morocco/abden-naser-aamar-325829
https://www.iaaf.org/athletes/ethiopia/gebrekidan-abadi-323754
https://www.iaaf.org/athletes/united-states/brandon-abasolo-364691
https://www.iaaf.org/athletes/ethiopia/tadu-abate-317050
https://www.iaaf.org/athletes/kuwait/abdul-redha-abbas-187218
https://www.iaaf.org/athletes/morocco/azzouzi-abdelaziz-221066
https://www.iaaf.org/athletes/chad/abdoulaye-abdelkarim-210297
https://www.iaaf.org/athletes/palestine/ihab-salama-abdellah-56172
https://www.iaaf.org/athletes/belgium/bashir-abdi-281105
https://www.iaaf.org/athletes/netherlands/mahadi-abdi-ali-350387
https://www.iaaf.org/athletes/italy/mohad-abdikadar-sheik-ali-263903
https://www.iaaf.org/athletes/united-states/abdihakem-abdirahman-137596
https://www.iaaf.org/athletes/ethiopia/ali-abdosh-242055
https://www.iaaf.org/athletes/chad/ahmat-abdou-daoud-243638
https://www.iaaf.org/athletes/republic-of-yemen/sofian-abdul-raggeb-136290
https://www.iaaf.org/athletes/ethiopia/shami-abdulahi-251403
https://www.iaaf.org/athletes/gatar/ahmad-hassan-abdullah-138182
https://www.iaaf.org/athletes/japan/hiroki-abe-310866
```





#### PACE CHART



MILE BEST	5K BEST / AVG MILE PACE	10K BEST / AVG MILE PACE	TEMPO AVG MILE PACE	HALF MARATHON BEST / AVG MILE PACE	MARATHON BEST / AVG MILE PACE	RECOVERY DAY Page
5:00	17:05 / <b>5:30</b>	35:45 / <b>5:45</b>	6:05	1:18:00 / <b>6:00</b>	2:44:00 / 6:15	7:00
5:30	18:45 / <b>6:00</b>	39:00 / <b>6:15</b>	6:35	1:25:00 / <b>6:30</b>	3:00:00 / 6:50	7:35
6:00	20:15 / <b>6:30</b>	42:00 / <b>6:45</b>	7:05	1:35:00 / <b>7:15</b>	3:15:00 / <b>7:25</b>	8:10
6:30	22:00 / <b>7:05</b>	45:45 / <b>7:20</b>	7:40	1:40:00 / <b>7:35</b>	3:30:00 / <b>8:00</b>	8:45
7:00	23:45 / <b>7:40</b>	49:00 / <b>7:55</b>	8:15	1:50:00 / <b>8:20</b>	3:45:00 / <b>8:35</b>	9:20
7:30	25:15 / <b>8:05</b>	52:30 / <b>8:25</b>	8:50	1:55:00 / <b>8:45</b>	4:00:00 / 9:10	9:55
8:00	27:00 / <b>8:40</b>	55:50 / <b>9:00</b>	9:25	2:05:00 / <b>9:30</b>	4:15:00 / <b>9:45</b>	10:30

## Next Steps

- Build prediction model
- Apply athletes' data to the training plan

The training strategy basically refer to the Half Marathon Pace Chart

## References

- •Thomas A. Severini. (2014). Analytic Methods in Sports: Using Mathematics and Statistics to Understand Data from Baseball, Football, Basketball, and Other Sports. Chapter 2.
- •S Tufféry. (2011). Data mining and statistics for decision making. 534 -538
- •Michael Bowles. (2015). Machine Learning in Python: Essential Techniques for Predictive Analysis. 122-124
- •Brian Hanley. Pacing profiles and pack running at the IAAF World Half Marathon Championships. (2015). https://pdfs.semanticscholar.org/819e/8db906d3673e46b6f609d1d77838f34ae1f1.pdf
- https://en.wikipedia.org/wiki/Naive\_Bayes\_classifier