

Дърво на решенията (Decision Tree). Пример.

Провеждане на игра на голф.

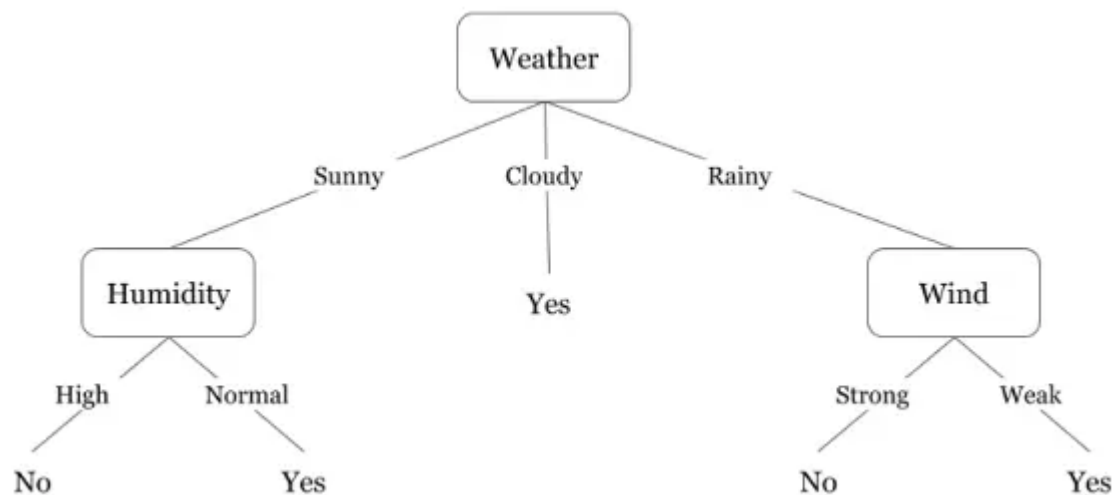
Модел Случайна гора (Random Forest)

<https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/>

Дървото на решенията се представя чрез диаграма при която коренът е отгоре и отбелязва началото на процеса. От корена се отделят две или повече направления, които завършват с възел. От всеки възел започва следващо разклонение и т.н докато се достигне до конкретно решение. Дървото на решенията представлява последователност от въпрос или условие. Всеки възел се интерпретира с някакъв въпрос или условие. В зависимост от отговорите се продължава към следващ възел или се достига до крайното решение.

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Cloudy	Hot	High	Weak	Yes
3	Sunny	Mild	Normal	Strong	Yes
4	Cloudy	Mild	High	Strong	Yes
5	Rainy	Mild	High	Strong	No
6	Rainy	Cool	Normal	Strong	No
7	Rainy	Mild	High	Weak	Yes
8	Sunny	Hot	High	Strong	No
9	Cloudy	Hot	Normal	Weak	Yes
10	Rainy	Mild	High	Strong	No

Да предположим, че искате да изберете дали да излезете навън и да играете или не. Можете да направите избор въз основа на времето. За това ето дърво на решенията:



Какво е времето ?		
Ако е слънчево и висока Влажност не играем. Ако е слънчево и нормална Влажност излизаме да играем.	Ако е облачно отиваме да играем.	Ако е дъждовно , питаме дали е ветровито. При силен вятър не играем. При слаб вятър започваме игра.

Използваме примера за игра на голф от лекцията за модела на наивния Бейс.

```

import numpy as np
data = np.asarray(((1,1,1,0),(1,1,1,1),(3,2,1,0), (3,3,0,0), (3,3,0,1), (2,3,0,1),
(1,2,1,0), (1,3,0,0), (3,2,0,0), (1,2,0,1),(2,2,1,1),(2,1,0,0),
(3,2,1,1)))
outcome = np.asarray((0,1,1,1,0,1,0,1,1,1,1,0))
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier()
today = np.asarray((3,1,0,0)) # array with 4 entries
today = today.reshape(1,-1) # га се обясни - it is an 1x4 array
model.fit(data, outcome)
z=model.predict(today)
print( 'z=',z)

```

Отново отговорът е z=1, при съответните условия за времето ще се проведе игра на голф.

Моделът дърво на решенията намира широки приложения в редица реални процеси и задачи. Това е основен модел при моделиране и анализ на данни !

Моделът Random Forest е съвкупност от дървета, прилагани в процеса на анализ на данни.

Прилагаме за предходния пример като използваме командите:

```
from sklearn.ensemble import RandomForestClassifier  
model = RandomForestClassifier()
```

Прилагаме модела Random Forest с класифициране на наблюденията от файла `framingham.csv`. Файлът е приложен.

Проф. Иван Иванов

Януари 2025