**Naïve Bayes’ Alogrithm**

***Scenario:*** Hãy tưởng tượng bạn đang cố gắng quyết định xem mình có nên đi chơi bóng đá vào cuối tuần này hay không, dựa trên các yếu tố như thời tiết và nhiệt độ.

Dữ liệu huấn luyện (trong quá khứ):

|  |  |  |  |
| --- | --- | --- | --- |
| Days | Weather | Temperature | Play or Not |
| 1 | Sunny | Warm | Yes |
| 2 | Cloudy | Warm | Yes |
| 3 | Cloudy | Cool | No |
| 4 | Rainy | Cold | No |
| 5 | Sunny | Warm | Yes |
| 6 | Cloudy | Cold | No |
| 7 | Sunny | Cool | Yes |
| 8 | Rainy | Warm | No |
| 9 | Rainy | Cold | No |
| 10 | Sunny | Warm | Yes |

**Mẫu dữ liệu mới (chúng ta muốn dự đoán):**

* **Thời tiết:** Nắng
* **Nhiệt độ:** Mát

**Câu hỏi:** Liệu chúng ta có nên chơi bóng đá không?

----------

***Step 1: Tính toán xác suất tiên nghiệm (Prior Probabilities)***

Count how many times “Yes” & “No”:

* “Yes”: 5 times
* “No”: 5 times
* Sum: 10

=> P( Play or not = Yes) = P (Play or not = No) =

***Step 2: Tính toán xác suất tiên nghiệm (Prior Probabilities)***

- Khi bóng đá = Có

+ P( Weather = Sunny|Yes) =

+ P(Weather = Cloudy|Yes) =

+ P(Weather = Rainy|Yes) = 0

+ P(Temperature = Warm|Yes) =

+ P(Temperature = Cool|Yes) =

**Khi Chơi bóng đá = Không:**

* **Thời tiết:**
  + P(Thời tieˆˊt = Na˘ˊng∣Khoˆng)=0/5=0
  + P(Thời tieˆˊt = Maˆy∣Khoˆng)=2/5
  + P(Thời tieˆˊt = Mưa∣Khoˆng)=3/5
* **Nhiệt độ:**
  + P(Nhiệt độ = Aˆˊm∣Khoˆng)=1/5
  + P(Nhiệt độ = Maˊt∣Khoˆng)=1/5
  + P(Nhiệt độ = Lạnh∣Khoˆng)=3/5

**Xử lý vấn đề tần suất bằng 0 (Zero-Frequency Problem) với Laplace Smoothing**

Như bạn thấy, có nhiều xác suất bằng 0. Nếu chúng ta dùng nguyên tắc nhân, bất kỳ xác suất bằng 0 nào cũng sẽ làm cho toàn bộ tích bằng 0, dẫn đến kết quả sai lệch. Để khắc phục, chúng ta sử dụng Laplace Smoothing (còn gọi là Add-1 Smoothing).

Công thức sau khi thêm Laplace Smoothing: P(Feature\_value∣Class)=(Count(Feature\_value,Class)+1)/(Count(Class)+Number\_of\_unique\_Feature\_values)

Áp dụng lại cho ví dụ của chúng ta:

**Khi Chơi bóng đá = Có (Count(Có) = 5):**

* Số lượng giá trị độc đáo của "Thời tiết" là 3 (Nắng, Mây, Mưa).
* Số lượng giá trị độc đáo của "Nhiệt độ" là 3 (Ấm, Mát, Lạnh).
* **Thời tiết:**
  + P(Thời tieˆˊt = Na˘ˊng∣Coˊ)=(4+1)/(5+3)=5/8
  + P(Thời tieˆˊt = Maˆy∣Coˊ)=(1+1)/(5+3)=2/8
  + P(Thời tieˆˊt = Mưa∣Coˊ)=(0+1)/(5+3)=1/8
* **Nhiệt độ:**
  + P(Nhiệt độ = Aˆˊm∣Coˊ)=(4+1)/(5+3)=5/8
  + P(Nhiệt độ = Maˊt∣Coˊ)=(1+1)/(5+3)=2/8
  + P(Nhiệt độ = Lạnh∣Coˊ)=(0+1)/(5+3)=1/8

**Khi Chơi bóng đá = Không (Count(Không) = 5):**

* **Thời tiết:**
  + P(Thời tieˆˊt = Na˘ˊng∣Khoˆng)=(0+1)/(5+3)=1/8
  + P(Thời tieˆˊt = Maˆy∣Khoˆng)=(2+1)/(5+3)=3/8
  + P(Thời tieˆˊt = Mưa∣Khoˆng)=(3+1)/(5+3)=4/8
* **Nhiệt độ:**
  + P(Nhiệt độ = Aˆˊm∣Khoˆng)=(1+1)/(5+3)=2/8
  + P(Nhiệt độ = Maˊt∣Khoˆng)=(1+1)/(5+3)=2/8
  + P(Nhiệt độ = Lạnh∣Khoˆng)=(3+1)/(5+3)=4/8

**Bước 3: Áp dụng Naive Bayes để dự đoán**

Chúng ta muốn tìm lớp có xác suất cao nhất cho mẫu dữ liệu mới: **Thời tiết = Nắng, Nhiệt độ = Mát**.

**Xác suất chơi bóng đá = Có:**

P(Na˘ˊng, Maˊt∣Coˊ)∗P(Coˊ) =P(Na˘ˊng∣Coˊ)∗P(Maˊt∣Coˊ)∗P(Coˊ) =(5/8)∗(2/8)∗0.5 =(10/64)∗0.5 =0.15625∗0.5=0.078125

**Xác suất chơi bóng đá = Không:**

P(Na˘ˊng, Maˊt∣Khoˆng)∗P(Khoˆng) =P(Na˘ˊng∣Khoˆng)∗P(Maˊt∣Khoˆng)∗P(Khoˆng) =(1/8)∗(2/8)∗0.5 =(2/64)∗0.5 =0.03125∗0.5=0.015625

**Bước 4: So sánh kết quả**

* P(Coˊ∣Na˘ˊng, Maˊt) tỉ lệ với 0.078125
* P(Khoˆng∣Na˘ˊng, Maˊt) tỉ lệ với 0.015625

Vì 0.078125>0.015625, thuật toán Naive Bayes dự đoán rằng **chúng ta nên chơi bóng đá** vào ngày hôm đó.