Lab 4

Lecturer: Nguyễn Hồ Duy Trí

Class: IS252.N11.KHCL

Student name: Hoàng Đình Hữu

Student ID: 20521384

Bài tập 3+4+5+6+7:

https://drive.google.com/drive/folders/1wbLf-9CfOnjAkYg1csBOPHLgU2EfZYEC?usp=sharing

Bài 1:

Thực hiện các yêu cầu đề cho:

Ảnh có chứa bàn

Mô tả được tạo tự động

1. Xác định tất cả những mâu thuẫn có thể có trong dữ liệu:

Trong dữ liệu bài tập trên không xảy ra các mâu thuẫn, không xuất hiện các dòng dữ liệu có thuộc tính giống nhau nhưng lại thuộc các phân lớp khác nhau.

1. Tính độ lợi thông tin (information gain) của thuộc tính và vẽ cây quyết định theo thuật toán ID3 cho dữ liệu:

* **Độ bất định của tập S: 8 dòng Thấp và 7 dòng Cao**

E(S) =

* **Độ lợi của thuộc tính Loại:**

* **Độ lợi của thuộc tính Số màu:**

* **Độ lợi của thuộc tính Kích thước:**

* **Độ lợi của thuộc tính Chất liệu:**

* Trong 4 thuộc tính xem xét, Chất liệu có độ lợi thông tin lớn nhất. Do đó, ta chọn thuộc tính này làm phép chia nhánh cho cây tại nút gốc.
* **Xét nhánh Nhựa PP :**

**Xét thuộc tính Loại :**

**Xét thuộc tính Số màu:**

**Xét thuộc tính Kích thước :**

* Với giá trị độ lợi dụng thông tin lớn hơn, cây sẽ tiếp tục được phân nhánh bằng thuộc tính Loại và nếu Loại = ‘Điều khiển’ thì nhánh Doanh số bán = ‘Cao’.
* **Xét nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Búp bê’**

**Xét thuộc tính Số màu:**

**Xét thuộc tính Kích thước:**

* **Xét nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Xếp hình’**

**Xét thuộc tính Số màu:**

**Xét thuộc tính Kích thước:**

*,*

* Với nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Búp bê’ thì ta chọn thuộc tính ‘Số màu’ làm phép chia nhánh.

Nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Búp Bê’:

* Có ‘Số màu’ = 3 thì nhánh luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.
* Có ‘Số màu’ = 5 thì nhánh luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.
* Có ‘Số màu’ = 7 thì nhánh luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.
* Với nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Xếp hình’ thì ta chọn thuộc tính ‘Kích thước’ làm phép chia nhánh.

Nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Xếp hình’:

* Có Kích thước = ‘To’ thì nhánh luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.
* Có Kích thước = ‘Nhỏ’ thì nhánh luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.
* **Xét nhánh Chất liệu = ‘Cao su’ với :**

**Xét thuộc tính Loại:**

**Xét thuộc tính Số màu:**

**Xét thuộc tính Kích thước:**

* Với nhánh Chất liệu = ‘Cao su’ thì cả 3 thuộc tính có độ lợi thông tin bằng nhau nên ta có thể chọn ngẫu nhiên một trong ba để tiếp tục chia nhánh, chọn thuộc tính Loại để làm phép chia nhánh.

Nhánh Chất liệu = ‘Cao su’ và:

* Có Loại = ‘Điều khiển’ thì nhánh luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.
* Có Loại = ‘Búp bê’ thì nhánh luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.
* **Xét nhánh Chất liệu = ‘Cao su’ và Loại = ‘Xếp hình’:**

**Xét thuộc tính Số màu:**

**Xét thuộc tính Kích thước:**

* Với nhánh Chất liệu = ‘Cao Su’ và Loại = ‘Xếp hình’, thì cả 2 thuộc tính đều có độ lợi thông tin bằng nhau nên ta có thể chọn ngẫu nhiên một trong ba thuộc tính để tiếp tục chia nhánh, ta chọn thuộc tính ‘Số màu’ làm phép chia nhánh.

Nhánh Chất liệu = ‘Cao su’ , Loại = ‘Xếp hình’

* Có Số màu = ‘5’ thì nhánh luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.
* Có Số màu = ‘7’ thì nhánh luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.

Vẽ cây quyết định:

Nhựa PP

Chất liệu

Cao Su

Loại

Loại

Loại

Búp bê

Điều khiển

Điều khiển

Xếp hình

Búp bê

Xếp hình

Số màu

Kích thước

Số màu

Thấp

Thấp

Cao

7

Nhỏ

To

7

5

3

5

Cao

Thấp

Thấp

Thấp

Cao

Cao

Cao

1. Tính giá trị chỉ số Gini (gini index) của các thuộc tính và vẽ cây quyết định theo thuật toán CART cho dữ liệu:

* **Gini của thuộc tính Loại:**
* **Gini của thuộc tính Số màu:**
* **Gini của thuộc tính Kích thước:**
* **Gini của thuộc tính Chất liệu:**

Thuộc tính Chất liệu có Gini thấp nhất nên ta chọn thuộc tính này làm phép chia nhánh cho cây tại nút gốc.

* **Xét nhánh Chất liệu = ‘Nhựa PP’**:

Gini của thuộc tính Loại:

Gini của thuộc tính Số màu:

Gini của thuộc tính Kích thước:

Nhánh Chất liệu = ‘Nhựa PP’, thuộc tính Loại có chỉ số Gini thấp nhất nên ta chọn thuộc tính này để tiếp tục chia nhánh.

Nhánh Chất liệu = ‘Nhựa PP’, Loại = ‘Điều khiển’ thì nhánh này luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.

Nhánh Chất liệu = ‘Nhựa PP, Loại = ‘Xếp hình’:

* Gini của thuộc tính Số màu:
* Gini của thuộc tính Kích thước:

Nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Xếp hình’:

* Có Kích thước = ‘Nhỏ’ thì nhánh này luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.
* Có Kích thước = ‘To’ thì nhánh này luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.

Nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Búp bê’:

* Gini của thuộc tính Số màu:
* Gini của thuộc tính Kích thước:

Nhánh Chất liệu = ‘Nhựa PP’ , Loại = ‘Búp bê’, chỉ số Gini của thuôc tính Số màu là thấp nhất nên ta lấy thuộc tính này để tiếp tục chia nhánh.

Nhánh Chất liệu = ‘Nhựa PP’ và Loại = ‘Búp bê’:

* Có Số màu = ‘3’ thì nhánh này luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.
* Có Số màu = ‘5’ thì nhánh này luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.
* Có Số màu = ‘7’ thì nhánh này luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.
* **Xét nhánh Chất liệu = ‘Cao su’**

Gini của thuộc tính Loại:

Gini của thuộc tính Số màu:

Gini của thuộc tính Kích thước:

Nhánh Chất liệu = ‘Cao su’ có cả 3 thuộc tính có chỉ số Gini bằng nhau nên ta có thể chọn ngẫu nhiên bất kì thuộc tính nào để tiếp tục chia nhánh, chọn thuộc tính Loại để tiếp tục chia nhánh.

Nhánh Chất liệu = ‘Cao su’:

* Có Loại = ‘Điều khiển’ thì nhánh này luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.
* Có Loại = ‘Búp bê’ thì nhánh này luôn có phân lớp Doanh số bán = ‘Thấp’ nên không cần xét tiếp.

Nhánh Chất liệu = ‘Cao su’ và Loại = ‘Xếp hình’:

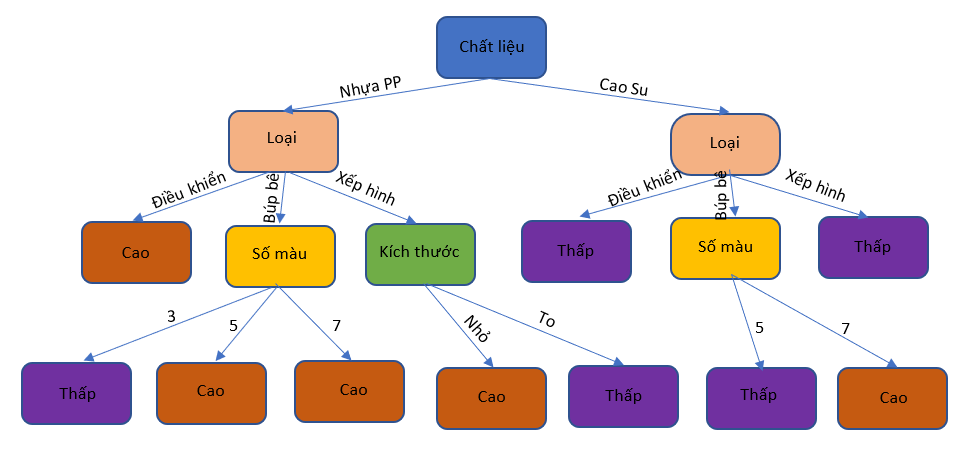
* Gini của thuộc tính Số màu:
* Gini của thuộc tính Kích thước:

Nhánh Chất liệu = ‘Cao Su’ và Loại = ‘Xếp hình’ có cả 2 thuộc tính có chỉ số Gini bằng nhau nên ta có thể lấy ngẫu nhiên một trong hai thuộc tính để tiếp tục chia nhánh, ta chọn thuộc tính Số màu để tiếp tục chia nhánh.

Nhánh Chất liệu = ‘Cao su’ , Loại = ‘Xếp hình’:

* Có Số màu = ‘5’ thì nhánh này luôn có phân lớp Doanh số bán = ’Thấp’ nên không cần xét tiếp.
* Có Số màu = ‘7’ thì nhành này luôn có phân lớp Doanh số bán = ‘Cao’ nên không cần xét tiếp.

Vẽ cây quyết định:



1. Sử dụng một trong hai cây quyết định ở trên để tiên đoán giá trị Doanh số bán của những sản phẩm sau:

Ảnh có chứa bàn

Mô tả được tạo tự động

Nếu Chất liệu = ‘Nhựa PP’, Loại = ‘Xếp hình’ và Kích thước = ‘To’ thì có Doanh số bán = ‘Thấp’.

Nếu Chất liệu = ‘Cao su’, Loại = ‘Điều khiển’ thì Doanh số bán = ‘Thấp’

1. Doanh số bán trên thực tế của các sản phẩm ở Yêu cầu d lần lượt là Thấp, Thấp, Cao. Lập ma trận nhầm lẫn, sau đó tính giá trị độ chính xác, độ phủ của mô hình/cây đã xây dựng:

|  |  |  |  |
| --- | --- | --- | --- |
| Lớp dự đoán mô hình | | | |
| Lớp trên thực tế |  | Thấp | Cao |
| Thấp | 2 | 0 |
| Cao | 1 | 0 |

Precision(M)

Recall(M) )

1. Xác suất không điều kiện của giá trị ‘Xếp hình’ trong tập dữ liệu:

P(Loại = Xếp hình)

1. Khi doanh số bán là ‘Thấp’, tính xác suất đó là những sản phẩm có chất liệu là ‘Cao su’:

P(Chất liệu = Cao su| Doanh số bán = Thấp)

1. Dựa theo định lí Bayes, viết công thức tính xác suất Doanh số ‘Cao’ của những sản phẩm thuộc loại ‘Điều khiển’:

1. Sử dụng thuật toán Naive Bayes và làm trơn Laplace để dự đoán giá trị Doanh số bán của những sản phẩm trong Yêu cầu d:

Áp dụng làm trơn Laplace ta có:

* P(Doanh số bán = Cao)
* P(Doanh số bán = Thấp)

Xét lần lượt từng dòng dữ liệu

Với dòng dữ liệu đầu tiên X ={Loại = ‘Búp bê’, Số màu=’3’,Kích thước =’To’, Chất liệu = ‘Cao su’}

* Xét trường hợp Doanh số bán = Cao:
* Xét trường hợp Doanh số bán = Thấp:

Với dòng dữ liệu đầu tiên có xác suất Doanh số bán = ‘Thấp’ lớn hơn nên ta có thể kết luận dòng dữ liệu đầu tiên được dự đoán vào phân lớp Doanh số bán = ‘Thấp’.

Với dòng dữ liệu thứ hai Y = {Loại = ‘Xếp hình’, Số màu=’5’, Kích thước=’To’, Chất lượng = ‘Nhựa PP’}

* Xét trường hợp Doanh số bán = Cao:
* Xét trường hợp Doanh số bán = Thấp:

Với dòng dữ liệu thứ hai có xác suất Doanh số bán = ‘Cao’ lớn hơn nên ta có thể kết luận dòng dữ liệu thứ hai được dự đoán vào phân lớp Doanh số bán = ‘Cao’.

Với dòng dữ liệu thứ ba Z= {Loại = ‘Điều khiển’, Số màu=’3’, Kích thước=’Vừa’, Chất lượng=‘Cao su’}

* Xét trường hợp Doanh số bán = Cao:
* Xét trường hợp Doanh số bán = Thấp:

Với dòng dữ liệu thứ ba có xác suất Doanh số bán = ‘Thấp’ lớn hơn nên ta có thể kết luận dòng dữ liệu thứ ba được dự đoán vào phân lớp Doanh số bán = ‘Thấp’.

1. VớI kết quả thu được và doanh số trên thực tế (Yêu cầu e), hãy lập ma trận nhầm lẫn, sau đó tính giá trị độ chính xác, độ phủ của thuật toán:

|  |  |  |  |
| --- | --- | --- | --- |
| Lớp dự đoán mô hình | | | |
| Lớp trên thực tế |  | Thấp | Cao |
| Thấp | 1 | 1 |
| Cao | 1 | 0 |

Precision(M)

Recall(M) )

1. So sánh kết quả từ thuật toán cây quyết định và Naive Bayes:

Từ ma trận nhầm lẫn và giá trị của độ chính xác, độ phủ trên dữ liệu kiểm tra thử được cho ở câu d, ta có thể kết luận được mô hình xây dựng dựa trên thuật toán Naive Bayes có độ chính xác thấp hơn mô hình xây dựng dựa trên cây quyết định ở câu a và câu b.

1. Sản phẩm mới của doanh nghiệp tung ra thị trường có thông tin như sau:

Ảnh có chứa bàn

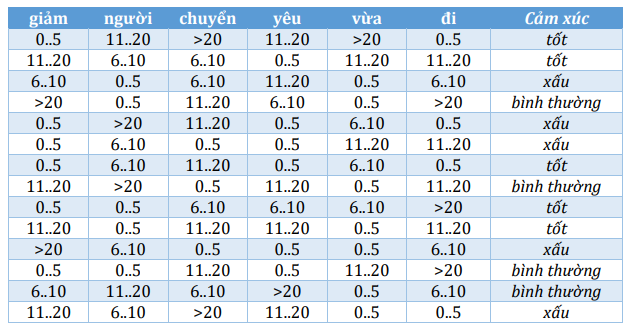
Mô tả được tạo tự động

T= {Loại = ‘Xếp hình, Số màu=’7’, Kích thước=’Nhỏ’, Chất lượng=‘Cao su’}

* Xét trường hợp Doanh số bán = Cao:
* Xét trường hợp Doanh số bán = Thấp:

Với dòng dữ liệu này có xác suất Doanh số bán = ‘Thấp’ lớn hơn nên ta có thể kết luận dòng dữ liệu này được dự đoán vào phân lớp Doanh số bán = ‘Thấp’.

Bài 2:



1. Xác định tất cả những mâu thuẫn có thể có trong dữ liệu:

Không có mâu thuẫn nào xuất hiện trong bộ dữ liệu này

1. Tính giá trị chỉ số Gini của các thuộc tính và vẽ cây quyết định theo thuật toán CART cho dữ liệu trên:

* **Gini của thuộc tính Giảm:**

* **Gini của thuộc tính Người:**

5

* **Gini của thuộc tính Chuyển:**

* **Gini của thuộc tính Yêu:**

* **Gini của thuộc tính Vừa:**

* **Gini của thuộc tính Đi:**

Thuộc tính Đi có giá trị Gini thấp nhất nên ta chọn thuộc tính này làm phép chia nhánh cho cây tại nút gốc.

* **Xét nhánh thuộc tính Đi = 0...5:**
* Gini thuộc tính Giảm với điều kiện Đi = 0…5:

* Gini thuộc tính Người với điều kiện Đi = 0…5:

* Gini thuộc tính Chuyểnvới điều kiện Đi = 0…5:

* Gini thuộc tính Yêu với điều kiện Đi = 0…5:

* Gini thuộc tính Vừa với điều kiện Đi = 0…5:

Nhánh Đi = 0…5 có 2 thuộc tính Người và Vừa có chỉ số Gini nhỏ nhất nên ta có thể chọn một trong hai thuộc tính để tiếp tục chia nhánh. Ta chọn thuộc tính Người để tiếp tục chia nhánh.

Xét nhánh Đi = 0…5:

* Có thuộc tính Người = ‘11…20’, với nhánh này luôn có phân lớp Cảm xúc = ‘Tốt’ nên không cần xét tiếp.
* Có thuộc tính Người = ‘>20’, với nhánh này luôn có phân lớp Cảm xúc = ‘Xấu’ nên không cần xét tiếp.

Xét nhánh Đi = 0…5 và Người= 6…10:

* Gini của thuộc tính Giảm:

* Gini của thuộc tính Chuyển:

* Gini của thuộc tính Yêu :

* Gini của thuộc tính Vừa :

Nhánh Đi = 0…5 và Người = 6…10 có tất cả các thuộc tính có chỉ số Gini bằng nhau ta có thể chọn bất kì một trong bốn thuộc tính,ta chọn thuộc tính Giảm để tiếp tục chia nhánh.

* **Xét nhánh Đi = 6...10:**
* Gini của thuộc tính Giảm:

* Gini của thuộc tính Người:

* Gini của thuộc tính Chuyển:

* Gini của thuộc tính Yêu:

* Gini của thuộc tính Vừa:

Nhánh Đi = 6…10 có 2 thuộc tính Người và Yêu có chỉ số Gini bằng nhau và nhỏ nhất nên ta có thể lấy một trong hai thuộc tính để tiếp tục chia nhánh, ta chọn thuộc tính Người để tiếp tục chia nhánh.

Xét nhánh Đi = 6…10:

* Có thuộc tính Người = 0…5, với nhánh này luôn có phân lớp Cảm xúc = ‘Xấu’ nên không cần xét tiếp.
* Có thuộc tính Người = 6…10, với nhánh này luôn có phân lớp Cảm xúc = ‘Xấu’ nên không cần xét tiếp.
* Có thuộc tính Người = 11...20, với nhánh này luôn có phân lớp Cảm xúc = ‘Bình thường’ nên không cần xét tiếp.
* **Xét nhánh Đi = 11..20:**
* Gini của thuộc tính Giảm:

* Gini của thuộc tính Người:

* Gini của thuộc tính Chuyển:

* Gini của thuộc tính Yêu:

* Gini của thuộc tính Vừa:

Nhánh Đi = 11...20 có 2 thuộc tính Người và Chuyển có chỉ số Gini bằng nhau và nhỏ nhất nên ta có thể lấy một trong hai thuộc tính để tiếp tục chia nhánh,ta chọn thuộc tính Người để tiếp tục chia nhánh.

Xét nhánh Đi = 11…20:

* Có thuộc tính Người = 0…5, với nhánh này luôn có phân lớp Cảm xúc= ‘Tốt’ nên không cần xét tiếp.
* Có thuộc tính Người= >20, với nhánh này luôn có phân lớp Cảm xúc= ‘Bình thường’ nên không cần xét tiếp.
* **Xét nhánh Đi = 11…20 và Người’ = 6…10:**
* Gini của thuộc tính Giảm:

* Gini thuộc tính Chuyển:

* Gini của thuộc tính Yêu:

* Gini của thuộc tính Vừa:

Nhánh Đi = 11…20 và Người = 6…10 có 2 thuộc tính Giảm và chuyển có chỉ số Gini bằng nhau và nhỏ nhất, ta có thể chọn một trong hai để tiếp tục phân nhánh, ta chọn thuộc tính Giảm để tiếp tục chia nhánh.

Xét nhánh Đi = 11..20, Người = 6…10:

* Có thuộc tính Giảm = 0…5, với nhánh này luôn có phân lớp Cảm xúc = ‘Xấu’ nên không cần xét tiếp.
* Có thuộc tính Giảm = 11…20, với nhánh này luôn có phân lớp Cảm xúc= ‘Tốt’ nên không cần xét tiếp.
* **Xét nhánh Đi = ‘>20’:**
* Gini của thuộc tính Giảm:

* Gini của thuộc tính Người:

* Gini của thuộc tính Chuyển:

* Gini của thuộc tính Yêu:

* Gini của thuộc tính Vừa:

Nhánh Đi = ‘>20’ có có 2 thuộc tính Chuyển và Vừa có chỉ số Gini bằng nhau và nhỏ nhất nên ta có thể chọn một trong hai thuộc tính để tiếp tục phân nhánh, ta chọn thuộc tính vừa để tiếp tục chia nhánh.

Xét nhánh Đi = ‘>20’:

* Có thuộc tính Vừa = 0…5, với nhánh này luôn có phân lớp Cảm xúc= ‘Bình thường’ nên không cần xét tiếp.
* Có thuộc tính Vừa= 6…10, với nhánh này luôn có phân lớp Cảm xúc= ‘Tốt’ nên không cần xét tiếp.
* Có thuộc tính Vừa =11…20, với nhánh này luôn có phân lớp Cảm xúc= ‘Bình thường’ nên không cần xét tiếp.

1. Sử dụng cây quyết định và thuật toán Naive Bayes để dự đoán cảm xúc của những trạng thái sau:

Ảnh có chứa văn bản, tường, trong nhà, ảnh chụp màn hình

Mô tả được tạo tự động

Sử dụng thuật toán Naïve Bayes:

Áp dụng làm trơn Laplace ta có

* P(Cảm xúc = Tốt) = =
* P(Cảm xúc = Xấu) = =
* P(Cảm xúc = Bình thường) = = 0,29

Với dòng dữ liệu X = {Giảm: 0..5, Người: 6..10, Chuyển: 0..5, Yêu: 11..20, Vừa: 6..10, Đi 0..5}

* Xét trường hợp Cảm xúc = Tốt:

* Xét trường hợp Cảm xúc = Xấu:
* Xét trường hợp Cảm xúc = Bình thường:

Dòng dữ liệu đầu tiên có xác suất Cảm xúc = ‘Xấu’ lớn nhất, vậy ta có thể kết luận dòng dữ liệu đầu tiên được dự đoán vào phân lớp Cảm xúc = ‘Xấu’.

Với dòng dữ liệu thứ hai Y = { Giảm: 0..5, Người: 0..5, Chuyển: 6..10, Yêu: 0..5, Vừa: 11..20, Đi >20}

* Xét trường hợp Cảm xúc=Tốt:
* Xét trường hợp Cảm xúc = Xấu:
* Xét trường hợp Cảm xúc = Bình thường:

Dòng dữ liệu thứ hai có xác suất Cảm xúc = ‘Tốt’ lớn nhất, vậy ta có thể kết luận dòng dữ liệu thứ hai được dự đoán vào phân lớp ‘cảm xúc’ = ‘Tốt’.

Với dòng dữ liệu thứ ba Z = { Giảm: 6...10, Người: 0..5, Chuyển: 11..20, Yêu: >20, Vừa: 6..10, Đi: 6..10}

* Xét trường hợp Cảm xúc = Tốt:
* Xét trường hợp Cảm xúc= Xấu:
* Xét trường hợp Cảm xúc = Bình thường:

Dòng dữ liệu thứ ba có xác suất Cảm xúc= ‘Bình thường’ lớn nhất, vậy ta có thể kết luận dòng dữ liệu thứ ba được dự đoán vào phân lớp Cảm xúc = ‘Bình thường’.

Với dòng dữ liệu thứ tư T = {Giảm: 6..10, Người: 11..20, Chuyển: 6..10, Yêu: 6..10, Vừa: >20, Đi: 0..5}

* Xét trường hợp Cảm xúc =Tốt:
* Xét trường hợp Cảm xúc=Xấu:
* Xét trường hợp Cảm xúc =Bình thường:

Dòng dữ liệu thứ tư có xác suất Cảm xúc = ‘Tốt’ lớn nhất, vậy ta có thể kết luận dòng dữ liệu thứ tư được dự đoán vào phân lớp Cảm xúc = ‘Tốt’.

1. Trên thực tế những trạng thái này lần lượt có cảm xúc là: xấu, tốt, bình thường. Hãy lập ma trận nhầm lẫn sau đó tính giá trị độ chính xác, độ phủ của cả hai phương pháp trên rồi so sánh chúng với nhau.

* CART:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lớp dự đoán mô hình | | | | |
| Lớp trên thực tế |  | Xấu | Tốt | Bình thường |
| Xấu | 0 | 1 | 0 |
| Tốt | 0 | 1 | 1 |
| Bình thường | 1 | 0 | 0 |

* NAIVE BAYES:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lớp dự đoán mô hình | | | | |
| Lớp trên thực tế |  | Xấu | Tốt | Bình thường |
| Xấu | 1 | 0 | 0 |
| Tốt | 0 | 2 | 0 |
| Bình thường | 1 | 0 | 1 |

Ta có thể thấy được thuật toán Naive Bayes có độ chính xác cao hơn CART

1. Nếu nắm bắt được cảm xúc của người dùng mạng xã hội thì sinh viên sẽ sử dụng chúng như thế nào?

* Dựa trên cảm xúc hiện tại của người dùng để cho ra các bài viết hoặc quảng cáo hay đơn giản là một đoạn phim để người dùng xem sao cho phù hợp với tâm trạng.
* Nếu nắm bắt được cảm xúc của người dùng thì sẽ giảm bớt đi ảnh hưởng tiêu cực của người dùng đó sử dụng vào mạng xã hội, từ đó ban hành các luật lệ về nội dung người dùng, cấm đăng tải mang tính chất tiêu cực hoặc bạo lực.