Bài 1 – Thực hành

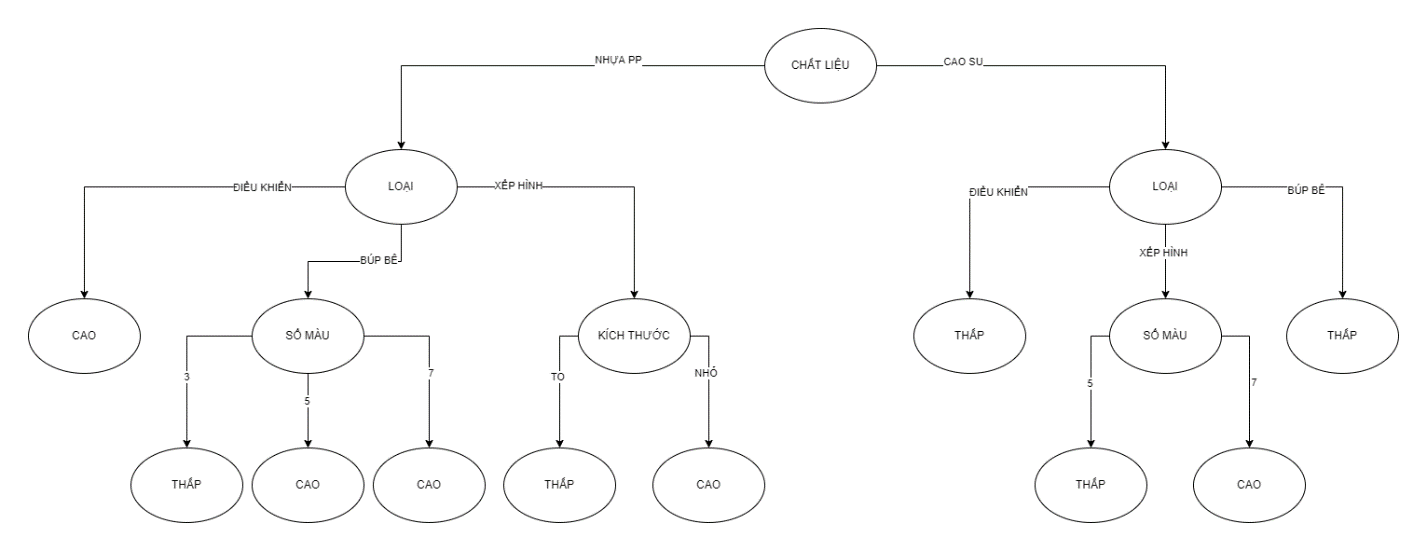
(Cơ bản) Một doanh nghiệp sản xuất đồ chơi cho trẻ em muốn dự đoán doanh số của các sản phẩm sắp đưa ra thị trường, họ thu thập những dữ liệu dưới đây:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Loại** | **Số màu** | **Kích thước** | **Chất liệu** | ***Doanh số bán*** |
| 1 | Điều khiển | 3 | Nhỏ | Nhựa PP | *Cao* |
| 2 | Xếp hình | 5 | Vừa | Cao su | *Thấp* |
| 3 | Xếp hình | 7 | To | Nhựa PP | *Thấp* |
| 4 | Điều khiển | 5 | Nhỏ | Cao su | *Thấp* |
| 5 | Búp bê | 3 | Vừa | Nhựa PP | *Thấp* |
| 6 | Điều khiển | 5 | Vừa | Nhựa PP | *Cao* |
| 7 | Búp bê | 5 | To | Nhựa PP | *Cao* |
| 8 | Điều khiển | 7 | Vừa | Cao su | *Thấp* |
| 9 | Xếp hình | 7 | To | Cao su | *Cao* |
| 10 | Xếp hình | 3 | To | Nhựa PP | *Thấp* |
| 11 | Búp bê | 3 | Nhỏ | Cao su | *Thấp* |
| 12 | Xếp hình | 3 | Nhỏ | Nhựa PP | *Cao* |
| 13 | Điều khiển | 5 | To | Cao su | *Thấp* |
| 14 | Búp bê | 5 | Vừa | Nhựa PP | *Cao* |
| 15 | Búp bê | 7 | To | Nhựa PP | *Cao* |

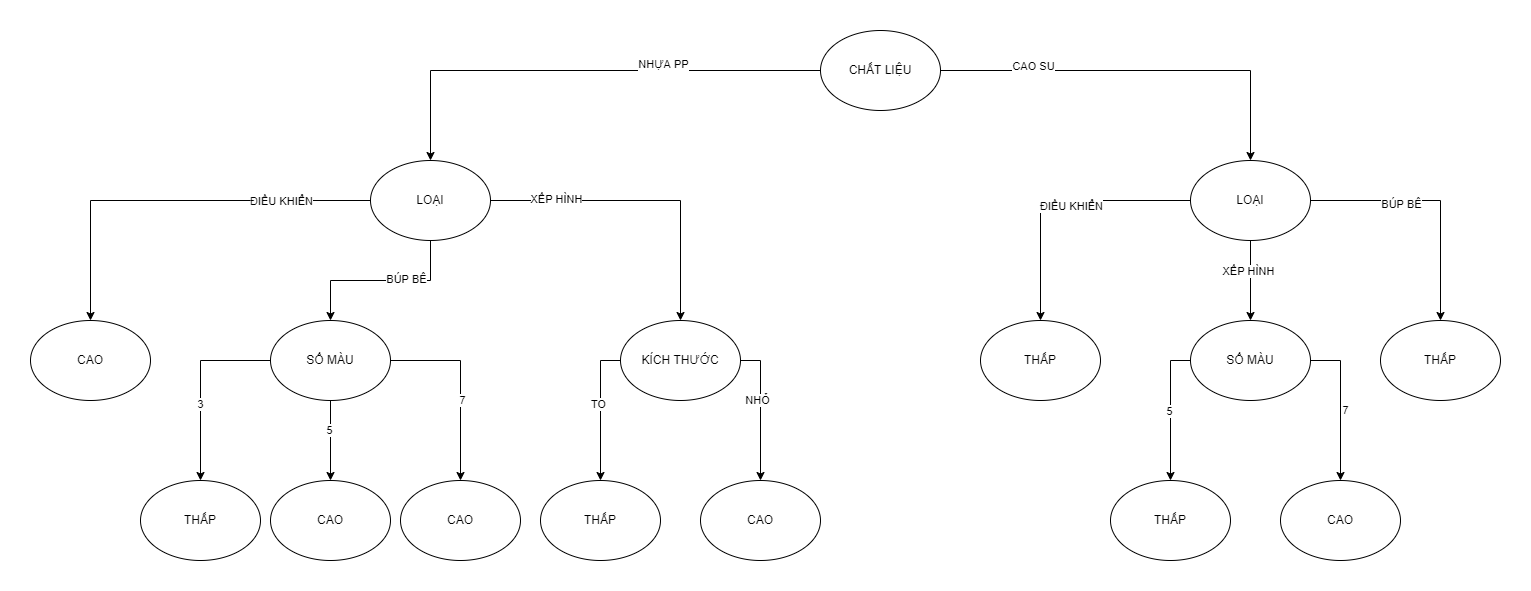
* + 1. Xác định tất cả những mâu thuẫn có thể có trong dữ liệu.
* Sau khi tiền xử lý dữ liệu cho phù hợp với thuật toán phân lớp, người ta thường rà soát lại để tìm ra những mâu thuẫn. Mâu thuẫn trong tập dữ liệu phân lớp là những dòng dữ liệu có giá trị thuộc tính giống nhau nhưng lại thuộc các phân lớp khác nhau. Trong dữ liệu bài tập trên không xảy ra trường hợp này.
  + 1. Tính giá trị độ lợi thông tin (information gain) của các thuộc tính và vẽ cây quyết

định theo thuật toán ID3 cho dữ liệu trên.

* Ban đầu tập dữ liệu S gồm 15 dòng dữ liệu, trong đó 7 dòng Doanh số bán *cao,* và 8 dòng doanh số bán **thấp**. Ta có Etropy (Độ bất định) của tập S như sau:
* ****
* Độ lợi thông tin của thuộc tính ***Loại***
* G(S, Loại) = 0,997 - =0,026
* Độ lợi thông tin của thuộc tính ***Số màu***
* G(S, Số màu) = 0,997 - 0,00667
* Độ lợi thông tin của thuộc tính ***Kích thước***
  + - G(S, Kích thước) = 0,997 - = 0,00667
* Độ lợi thông tin của thuộc tính ***Chất liệu***
* = 0,997 - -
* Trong 4 thuộc tính ta xét, thuộc tính ***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 trường hợp 1**: Nhánh chất liệu = “Nhựa PP”
* Độ lợi của thuộc tính ***loại*** với điều kiện nhánh chất liệu = “Nhựa PP”
* ;
* ;
* ;
* Độ lợi của thuộc tính ***số màu*** với điều kiện nhánh chất liệu = “Nhựa PP”
* Độ lợi của thuộc tính ***kích thước*** với điều kiện nhánh chất liệu = “Nhựa PP”
* Vậy nhánh chất liệu = “Nhựa PP” có thuộc tính ***loại*** có độ lợi thông tin lớn nhất nên ta lấy thuộc tính ***loại*** tiếp tục chia nhánh
* **Xét trường hợp 1.1:** Nhánh chất liệu = “Nhựa PP” và Loại = “Điều khiển”, với nhánh này luôn có phân lớp là Doanh số bán = “Cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.2:** Nhánh chất liệu = “Nhựa PP” và Loại = “Xếp hình”
* Độ lợi của thuộc tính ***số màu*** với điều kiện nhánh chất liệu = “Nhựa PP” và Loại = “Xếp hình”
* ;
* = 0,918 -
* Độ lợi của thuộc tính ***kích thước*** với điều kiện nhánh chất liệu = “Nhựa PP” và Loại = “Xếp hình”
* ;
* = 0,918 -
* Vậy nhánh chất liệu = “Nhựa PP” và Loại = “Xếp hình” có thuộc tính ***kích thước*** có độ lợi thông tin lớn nhất nên ta lấy thuộc tính ***kích thước*** tiếp tục chia nhánh.
* **Xét trường hợp 1.2.1:** Nhánh chất liệu = “Nhựa PP”, Loại = “Xếp hình”, Kích thước = “To” với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.2.2:** Nhánh chất liệu = “Nhựa PP”, Loại = “Xếp hình”, Kích thước = “Nhỏ” với nhánh này luôn có phân lớp là Doanh số bán = “Cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.3:** Nhánh chất liệu = “Nhựa PP” và Loại = “Búp bê”
* Độ lợi của thuộc tính ***số màu*** với điều kiện nhánh chất liệu = “Nhựa PP” và Loại = “Búp bê”
* ;
* = 0,8113 -
* Độ lợi của thuộc tính ***kích thước*** với điều kiện nhánh chất liệu = “Nhựa PP” và Loại = “Búp bê”
* ;
* = 0,8113
* Vậy nhánh chất liệu = “Nhựa PP” và Loại = “Búp bê” có thuộc tính ***Số màu*** có độ lợi thông tin lớn nhất nên ta lấy thuộc tính ***Số màu*** tiếp tục chia nhánh.
* **Xét trường hợp 1.3.1:** Nhánh chất liệu = “Nhựa PP”, Loại = “Búp bê”, Số màu= “3” với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.3.2:** Nhánh chất liệu = “Nhựa PP”, Loại = “Búp bê”, Số màu= “5” với nhánh này luôn có phân lớp là Doanh số bán = “cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.3.3:** Nhánh chất liệu = “Nhựa PP”, Loại = “Búp bê”, Số màu= “7” với nhánh này luôn có phân lớp là Doanh số bán = “cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 2**: Nhánh chất liệu = “Cao su”
* Độ lợi của thuộc tính ***loại*** với điều kiện nhánh chất liệu = “Cao su”
* ;
* ;
* ;
* Độ lợi của thuộc tính ***số màu*** với điều kiện nhánh chất liệu = “Cao su”
* ;
* ;
* ;
* Độ lợi của thuộc tính ***kích thước*** với điều kiện nhánh chất liệu = “Cao su”
* ;
* ;
* ;
* Vậy nhánh chất liệu = “Cao su” có 3 thuộc tính loại, số màu, kích thướccó độ lợi thông tin bằng nhau nên ta lấy bất kì thuộc tính nào tiếp tục chia nhánh, vậy bài này lấy thuộc tính ***loại*** tiếp tục chia nhánh.
* **Xét trường hợp 2.1:** Nhánh chất liệu = “Cao su”, Loại = “Điều khiển”với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 2.2:** Nhánh chất liệu = “Cao su”, Loại = “Búp bê”với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 2.3:** Nhánh chất liệu = “Cao su” và Loại = “Xếp hình”
* Độ lợi của thuộc tính ***số màu*** với điều kiện nhánh chất liệu = “Cao su” và Loại = “Xếp hình”
* Độ lợi của thuộc tính ***kích thước*** với điều kiện nhánh chất liệu = “Cao su” và Loại = “Xếp hình”
* Vậy nhánh chất liệu = “Cao su” và Loại = “Xếp hình”” có 2 số màu và kích thướccó độ lợi thông tin bằng nhau nên ta lấy bất kì thuộc tính nào tiếp tục chia nhánh, vậy bài này lấy thuộc tính ***số màu*** tiếp tục chia nhánh.
* **Xét trường hợp 2.3.1:** Nhánh chất liệu = “Cao su”, Loại = “Xếp hình”, Số màu= “5” với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 2.3.2:** Nhánh chất liệu = “Cao su”, Loại = “Xếp hình”, Số màu= “7” với nhánh này luôn có phân lớp là Doanh số bán = “cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* Ta có cây quyết định được xây dựng là như hình sau đây:



* + 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 trên.
* Chỉ số Gini thuộc tính **Loại**
* Chỉ số Gini thuộc tính **Số màu**
* Chỉ số Gini thuộc tính **Kích thước**
* Chỉ số Gini thuộc tính **Chất liệu**
* Trong 4 thuộc tính ta xét, thuộc tính ***chất liệu*** có chỉ số Gini nhỏ nhất. Do đó, ta chọn thuộc tính này làm phép chia nhánh cho cây tại nút gốc.
* **Xét trường hợp 1:** Nhánh chất liệu = “Nhựa PP”
* Chỉ số Gini thuộc tính **Loại** với điều kiến nhánh chất liệu = “Nhựa PP”
* Chỉ số Gini thuộc tính **Số màu** với điều kiến nhánh chất liệu = “Nhựa PP”
* ,5
* Chỉ số Gini thuộc tính **Kích thước** với điều kiến nhánh chất liệu = “Nhựa PP”
* Vậy nhánh chất liệu = “Nhựa PP” có thuộc tính ***loại*** có chỉ số Gini nhỏ nhất nên ta lấy thuộc tính ***loại*** tiếp tục chia nhánh.
* **Xét trường hợp 1.1:** Nhánh chất liệu = “Nhựa PP” và Loại = “Điều khiển”, với nhánh này luôn có phân lớp là Doanh số bán = “Cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.2:** Nhánh chất liệu = “Nhựa PP” và Loại = “Xếp hình”
* Chỉ số Gini thuộc tính **Số màu** với điều kiến nhánh chất liệu = “Nhựa PP” và Loại = “Xếp hình”
* ,5
* Chỉ số Gini thuộc tính **Kích thước** với điều kiến nhánh chất liệu = “Nhựa PP” và Loại = “Xếp hình”
* **Xét trường hợp 1.2.1:** Nhánh chất liệu = “Nhựa PP”, Loại = “Xếp hình”, Kích thước = “To” với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.2.2:** Nhánh chất liệu = “Nhựa PP”, Loại = “Xếp hình”, Kích thước = “Nhỏ” với nhánh này luôn có phân lớp là Doanh số bán = “Cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.3:** Nhánh chất liệu = “Nhựa PP” và Loại = “Búp bê”
* Chỉ số Gini thuộc tính **Số màu** với điều kiến nhánh chất liệu = “Nhựa PP” và Loại = “Búp bê”
* Chỉ số Gini thuộc tính **Kích thước** với điều kiến nhánh chất liệu = “Nhựa PP” và Loại = “Búp bê”
* Vậy nhánh chất liệu = “Nhựa PP” và Loại = “Búp bê” có thuộc tính ***Số màu*** có chỉ số gini nhỏ nhất nên ta lấy thuộc tính ***Số màu*** tiếp tục chia nhánh.
* **Xét trường hợp 1.3.1:** Nhánh chất liệu = “Nhựa PP”, Loại = “Búp bê”, Số màu= “3” với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.3.2:** Nhánh chất liệu = “Nhựa PP”, Loại = “Búp bê”, Số màu= “5” với nhánh này luôn có phân lớp là Doanh số bán = “cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 1.3.3:** Nhánh chất liệu = “Nhựa PP”, Loại = “Búp bê”, Số màu= “7” với nhánh này luôn có phân lớp là Doanh số bán = “cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 2**: Nhánh chất liệu = “Cao su”
* Chỉ số Gini thuộc tính **Loại** với điều kiến nhánh chất liệu = “Cao su”
* Chỉ số Gini thuộc tính Số màu với điều kiện nhánh chất liệu = “Cao su”
* Chỉ số Gini thuộc tính **Kích thước** với điều kiến nhánh chất liệu = “Cao su”
* Vậy nhánh chất liệu = “Cao su” có 3 thuộc tính loại, số màu, kích thướccó chỉ số gini bằng nhau nên ta lấy bất kì thuộc tính nào tiếp tục chia nhánh, vậy bài này lấy thuộc tính ***loại*** tiếp tục chia nhánh.
* **Xét trường hợp 2.1:** Nhánh chất liệu = “Cao su”, Loại = “Điều khiển”với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 2.2:** Nhánh chất liệu = “Cao su”, Loại = “Búp bê”với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 2.3:** Nhánh chất liệu = “Cao su” và Loại = “Xếp hình”
* Chỉ số Gini thuộc tính **Số màu** với điều kiến nhánh chất liệu = “Cao su” và Loại = “Xếp hình”
* Chỉ số Gini thuộc tính **Kích thước** với điều kiến nhánh chất liệu = “Cao su” và Loại = “Xếp hình”
* Vậy nhánh chất liệu = “Cao su” và Loại = “Xếp hình”” có 2 số màu và kích thướccó chỉ số gini bằng nhau nên ta lấy bất kì thuộc tính nào tiếp tục chia nhánh, vậy bài này lấy thuộc tính ***số màu*** tiếp tục chia nhánh.
* **Xét trường hợp 2.3.1:** Nhánh chất liệu = “Cao su”, Loại = “Xếp hình”, Số màu= “5” với nhánh này luôn có phân lớp là Doanh số bán = “thấp”, vì vậy nhánh này nút lá không cần xét tiếp.
* **Xét trường hợp 2.3.2:** Nhánh chất liệu = “Cao su”, Loại = “Xếp hình”, Số màu= “7” với nhánh này luôn có phân lớp là Doanh số bán = “cao”, vì vậy nhánh này nút lá không cần xét tiếp.
* Ta có cây quyết định được xây dựng là như hình sau đây:



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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Loại** | **Số màu** | **Kích thước** | **Chất liệu** | **Doanh số bán** |
| Búp bê | 3 | To | Cao su | Thấp |
| Xếp hình | 5 | To | Nhựa PP | Thấp |
| Điều khiển | 3 | Vừa | Cao su | Thấp |

* + 1. Doanh số bán trên thực tế của các sản phẩm ở Yêu cầu d lần lượt là *Thấp, Thấp, Cao*. Hãy 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 |

* + 1. Xác suất không điều kiện của giá trị ‘Xếp hình’ trong tập dữ liệu là bao nhiêu?
* P(Loại = Xếp hình) =
  + 1. Khi doanh số bán là ‘Thấp’, hãy 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 lý Bayes, hãy viết công thức tính xác suất Doanh số ‘Cao’ của những sản phẩm thuộc loại ‘Điều khiển’.
* P(Doanh số bán = Cao|Loại = Điều khiển) = = = 0,4
  + 1. Sử dụng thuật toán Naïve Bayes và làm trơn Laplace để dự đoán giá trị Doanh số bán của những sản phẩm trong Yêu cầu d.

Áp dụng qui tắc làm trơn Laplace ta có

* P(Doanh số bán = Cao) = = 8/17
* P(Doanh số bán = Thấp) = = 9/17 0,5294
* Với tập hồ sơ X = {Loại = ‘Búp bê’, Số màu = ‘3’, Kích thước = ‘To’, Chất liệu = ‘Cao su’}
* Xét trường hợp doanh số bán = Cao
  + - P(1)=P(Loại = Búp bê| Doanh số bán = Cao) = 
    - P(2)=P(Số màu = 3| Doanh số bán = Cao) = 
    - P(3)=P(Kích thước = To| Doanh số bán = Cao) = 
    - P(4)=P(Chất liệu =Cao su| Doanh số bán = Cao) = 
    - P(Doanh số bán = Cao) \*P(X|Doanh số bán = Cao) = P(Doanh số bán = Cao)\*P(1)\*P(2)\*P(3)\*P(4) = 0,4706\*0,4\*0,3\*0,4\*
* Xét trường hợp doanh số bán = Thấp
  + - P(5)=P(Loại = Búp bê| Doanh số bán = Thấp) = 
    - P(6)=P(Số màu = 3| Doanh số bán = Thấp) = 
    - P(7)=P(Kích thước = To| Doanh số bán = Thấp) = 
    - P(8)=P(Chất liệu =Cao su| Doanh số bán = Thấp) = 
    - P(Doanh số bán = Thấp) \*P(X|Doanh số bán = Thấp) = P(Doanh số bán = Thấp)\*P(5)\*P(6)\*P(7)\*P(8) = 0,5294\*0,2727\*0,3636\*0,3636\*0,6
* Như vậy hồ sơ đầu tiên có xác suất xảy ra Doanh số bán = ‘Thấp’ lớn hơn, vậy ta có thể kết luận dòng dữ liệu đầu tiên được dự đoán vào phân lớp Doanh số bán = ‘Thấp’
* Với tập hồ sơ Y = {Loại = ‘Xếp hình’, Số màu = ‘5’, Kích thước = ‘To’, Chất liệu = ‘Nhựa PP’}
* Xét trường hợp doanh số bán = Cao
  + - P(9)=P(Loại = Xếp hình| Doanh số bán = Cao) = 
    - P(10)=P(Số màu = 5| Doanh số bán = Cao) = 
    - P(3)=P(Kích thước = To| Doanh số bán = Cao) = 
    - P(11)=P(Chất liệu =Nhựa PP| Doanh số bán = Cao) = 
    - P(Doanh số bán = Cao) \*P(Y|Doanh số bán = Cao)= P(Doanh số bán = Cao)\*P(9)\*P(10)\*P(3)\*P(11) = 0,4706\*0,3\*0,4\*0,4\*0,778 0,0176
* Xét trường hợp doanh số bán = Thấp
  + - P(12)=P(Loại = Xếp hình| Doanh số bán = Thấp) = 
    - P(13)=P(Số màu = 5| Doanh số bán = Thấp) = 
    - P(7)=P(Kích thước = To| Doanh số bán = Thấp) = 
    - P(14)=P(Chất liệu =Nhựa PP| Doanh số bán = Thấp) = 
    - P(Doanh số bán = Thấp) \*P(Y|Doanh số bán = Thấp) = P(Doanh số bán = Thấp)\*P(12)\*P(13)\*P(7)\*P(14) = 0,5294\*0,364\*0,364\*0,364\*0,4
* Như vậy hồ sơ đầu tiên có xác suất xảy ra Doanh số bán = ‘Cao’ lớn hơn, vậy ta có thể kết luận dòng dữ liệu đầu tiên được dự đoán vào phân lớp Doanh số bán = ‘Cao’
* Với tập hồ sơ Z = {Loại = ‘Điều khiển’, Số màu = ‘3’, Kích thước = ‘Vừa’, Chất liệu = ‘Cao su’}
* Xét trường hợp doanh số bán = Cao
  + - P(15)=P(Loại = Điều khiển| Doanh số bán = Cao) = 
    - P(2)=P(Số màu = 3| Doanh số bán = Cao) = 
    - P(16)=P(Kích thước = Vừa| Doanh số bán = Cao) = 
    - P(4)=P(Chất liệu =Cao su| Doanh số bán = Cao) = 
    - P(Doanh số bán = Cao) \*P(Z|Doanh số bán = Cao)= P(Doanh số bán = Cao)\*P(15)\*P(2)\*P(16)\*P(4) = 0,4706\*0,3\*0,3\*0,3\*
* Xét trường hợp doanh số bán = Thấp
  + - P(17)=P(Loại = Điều khiển| Doanh số bán = Thấp) = 
    - P(6)=P(Số màu = 3| Doanh số bán = Thấp) = 
    - P(18)=P(Kích thước = Vừa| Doanh số bán = Thấp) = 
    - P(8)=P(Chất liệu =Cao su| Doanh số bán = Thấp) = 
    - P(Doanh số bán = Thấp) \*P(Z|Doanh số bán = Thấp) = P(Doanh số bán = Thấp)\*P(17)\*P(6)\*P(18)\*P(8) = 0,5294\*0,364\*0,364\*0,364\*0,6
* Như vậy hồ sơ đầu tiên có xác suất xảy ra Doanh số bán = ‘Thấp’ lớn hơn, vậy ta có thể kết luận dòng dữ liệu đầu tiên 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 |

* Theo công thức ta có độ chính xác là:
  + 1. So sánh kết quả từ thuật toán cây quyết định và Naïve Bayes.
* Từ ma trận nhầm lẫn và giá trị của độ chính xác, độ phủ trên dữ liệu kiểm thử được cho ở câu 4, ta có kết luận được mô hình được xây dựng bởi thuật toán Naïve Bayes có độ chính xác thấp hơn mô hình cây quyết định được xây dựng ở câu a,b.
  + 1. Sản phẩm mới của doanh nghiệp dự định tung ra thị trường có thông tin như sau:

|  |  |  |  |
| --- | --- | --- | --- |
| **Loại** | **Số màu** | **Kích thước** | **Chất liệu** |
| Xếp hình | 7 | Nhỏ | Cao su |

Hãy sử dụng các mô hình đã xây dựng được để dự đoán Doanh số bán của công ty với sản phẩm này.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Loại** | **Số màu** | **Kích thước** | **Chất liệu** | **Doanh số bán** |
| Xếp hình | 7 | Nhỏ | Cao su | Cao |

* X = {Loại = Xếp hình, Số màu = 7, Kích thước = Nhỏ, Chất liệu = Cao su}
* Xét trường hợp doanh số bán = Cao
  + - P(Loại = Xếp hình| Doanh số bán = Cao) =
    - P(Số màu = 7| Doanh số bán = Cao) = 
    - P(Kích thước = Nhỏ| Doanh số bán = Cao) = 
    - P(Chất liệu =Cao su| Doanh số bán = Cao) = 
    - P(Doanh số bán = Cao) \*P(X|Doanh số bán = Cao) = P(Doanh số bán = Cao)\* P(Loại = Xếp hình| Doanh số bán = Cao)\* P(Số màu = 7| Doanh số bán = Cao)\* P(Kích thước = Nhỏ| Doanh số bán = Cao)\* P(Chất liệu =Cao su| Doanh số bán = Cao) = 0,4706\*0,3\*0,3\*0,3\*
* Xét trường hợp doanh số bán = Thấp
* P(Loại = Xếp hình| Doanh số bán = Thấp) = 
* P(Số màu = 7| Doanh số bán = Thấp) = 
* P(Kích thước = Nhỏ| Doanh số bán = Thấp) = 
* P(Chất liệu =Cao su| Doanh số bán = Thấp) = 
* P(Doanh số bán = Thấp) \*P(X|Doanh số bán = Thấp) = P(Doanh số bán = Thấp)\* P(Loại = Xếp hình| Doanh số bán = Thấp)\*P(Số màu = 7| Doanh số bán = Thấp)\* P(Kích thước = Nhỏ| Doanh số bán = Thấp)\* P(Chất liệu =Cao su| Doanh số bán = Thấp)= 0,5294\*0,364\*0,273\*0,273\*0,6
* Như vậy hồ sơ đầu tiên có xác suất xảy ra Doanh số bán = ‘Thấp’ lớn hơn, vậy ta có thể kết luận dòng dữ liệu đầu tiên được dự đoán vào phân lớp Doanh số bán = ‘Thấp”