A picture containing text, electronics

Description automatically generatedA picture containing text, businesscard

Description automatically generated

1. – CART

* Trường hợp Gini cho Temperature khi outlook = sunny

|  |  |  |  |
| --- | --- | --- | --- |
| Temperature | Yes | No | # Instances |
| Hot | 0 | 2 | 2 |
| Cool | 1 | 1 | 1 |
| mild | 1 | 1 | 2 |

𝐺𝑖𝑛𝑖(outlook = sunny & temperature = hot) = 1 − ( 0/ 2 )^2 − ( 2 /2 )^2 = 0

𝐺𝑖𝑛𝑖(outlook = sunny &temperature = cool) = 1 − ( 1/ 1 )^2 − ( 0 /1 )^2 = 0

𝐺𝑖𝑛𝑖(outlook = sunny & temperature = mild) = 1 − ( 1 /2 )^2 − ( 1 /2 )^2 = 0.5

𝐺𝑖𝑛𝑖(outlook = sunny & temperature) = 2/ 5 ∗ 0 + 1 /5 ∗ 0 + 2/ 5 ∗ 0.5 = 0.2

* Trường hợp Gini cho Humidity khi outlook = sunny

|  |  |  |  |
| --- | --- | --- | --- |
| Humidity | Yes | No | #Instances |
| High | 0 | 3 | 3 |
| Normal | 2 | 0 | 2 |

𝐺𝑖𝑛𝑖(outlook = sunny & humidity = high) = 1 − ( 0/ 3 )^2 − ( 3/ 3 )^2 = 0

𝐺𝑖𝑛𝑖(outlook = sunny &humidity = normal) = 1 − ( 2/ 2 )^2 − ( 0 /2 )^2 = 0

𝐺𝑖𝑛𝑖(outlook = sunny & humidity) = 3 /5 ∗ 0 + 2 /5 ∗ 0 = 0

* Trường hợp Gini cho Wind khi outlook = sunny

|  |  |  |  |
| --- | --- | --- | --- |
| Wind | Yes | No | # Instances |
| Weak | 1 | 2 | 3 |
| Strong | 1 | 1 | 2 |

𝐺𝑖𝑛𝑖(outlook = sunny & wind = weak) = 1 − ( 1 /3 ^2 − ( 2/ 3 )^2 = 0.44

𝐺𝑖𝑛𝑖(outlook = sunny &wind = strong) = 1 − ( 1/2 )^2 − ( 1/ 2 )^2 = 0.5

𝐺𝑖𝑛𝑖(outlook = sunny & wind) = 3 /5 ∗ 0.44 + 2/ 5 ∗ 0.5 = 0.466

|  |  |
| --- | --- |
| Features | Gini |
| Temperature | 0.2 |
| Humidity | 0 |
| wind | 0.466 |

Ta thấy Humidity có giá trị nhỏ nhất vì thế node tiếp theo là Humidity

rainfall

overcast

sunny

?

?

?

?

humidity

outlook

Trường hợp ocercast quyết đinh luôn là YES

outlook

humidity

?

?

?

YES

sunny

overcast

rainfall

Trường hợp Humidity

Humidity = high luôn quyết định là No

Humidity = Normal luôn quyết định là Yes

outlook

humidity

?

YES

sunny

overcast

rainfall

NO

YES

high

Normal

Trường hợp outlook = rainfall

𝐺𝑖𝑛𝑖(outlook = rainfall & temp = Cool) = 1 − ( 1 /2 )^2 − ( 1 /2 )^2 = 0.5

𝐺𝑖𝑛𝑖(outlook = rainfall & temp = Mild) = 1 − ( 2/ 3 ) ^2 − ( 1 /3 ) ^2 = 0.444

𝐺𝑖𝑛𝑖(outlook = rainfall & temp) = 2 /5 ∗ 0.5 + 3 /5 ∗ 0.444 = 0.466

𝐺𝑖𝑛𝑖(outlook = rainfall and humidity = high) = 1 − ( 1 /2 ) ^2 − ( 1/ 2 ) ^2 = 0.5

𝐺𝑖𝑛𝑖(outlook = rainfall and humidity = normal) = 1 − ( 2/ 3 )^ 2 − ( 1 /3 )^ 2 = 0.444

𝐺𝑖𝑛𝑖(outlook = rainfall & temp) = 2 5 ∗ 0.5 + 3 5 ∗ 0.444 = 0.466

𝐺𝑖𝑛𝑖(outlook = rainfall & wind = weak) = 1 − ( 3/ 3 )^2 − ( 0/ 3 )^2 = 0

𝐺𝑖𝑛𝑖(outlook = rainfall & wind = strong) = 1 − ( 0 /2 )^ 2 − ( 2 /2 )^2 = 0

𝐺𝑖𝑛𝑖(outlook = rainfall & and wind) = 3 /5 ∗ 0 + 2 /5 ∗ 0 = 0

outlook

humidity

wind

YES

sunny

overcast

rainfall

NO

YES

Normal

High

YES

NO

ID3

𝐺𝑎𝑖𝑛(𝑂𝑢𝑡𝑙𝑜𝑜𝑘) = 0.94 − 0.693 = 0.247

𝐺𝑎𝑖𝑛(𝑇𝑒𝑚𝑝𝑒𝑟𝑎𝑡𝑢𝑟𝑒) = 0.94 − 0.911 = 0.029

𝐺𝑎𝑖𝑛(𝐻𝑢𝑚𝑖𝑑𝑖𝑡𝑦) = 0.94 − 0.788 = 0.152

𝐺𝑎𝑖𝑛(𝑊𝑖𝑛𝑑𝑦) = 0.94 − 0.892 = 0.048

𝐺𝑎𝑖𝑛(𝐻𝑢𝑚𝑖𝑑𝑖𝑡𝑦 & 𝑜𝑢𝑡𝑙𝑜𝑜𝑘 = 𝑠𝑢𝑛𝑛𝑦) = 0.971 − 0 = 0.971

𝐺𝑎𝑖𝑛(𝑊𝑖𝑛𝑑𝑦 & 𝑜𝑢𝑡𝑙𝑜𝑜𝑘 = 𝑠𝑢𝑛𝑛𝑦) = 0.971 − 0.951 = 0.020

𝐺𝑎𝑖𝑛(𝑇𝑒𝑚𝑝𝑒𝑟𝑎𝑡𝑢𝑟𝑒 & 𝑜𝑢𝑡𝑙𝑜𝑜𝑘 = 𝑠𝑢𝑛𝑛𝑦) = 0.971 − 0.4 = 0.571

rainfall

overcast

sunny

?

humidity

outlook

YES

Normal

High

YES

NO

𝐺𝑎𝑖𝑛(𝐻𝑢𝑚𝑖𝑑𝑖𝑡𝑦 & Outlook = Rainfall) = 0.971 − 0.951 = 0.020

𝐺𝑎𝑖𝑛(𝑊𝑖𝑛𝑑𝑦 & Outlook = Rainfall) = 0.971 − 0 = 0.971

𝐺𝑎𝑖𝑛(𝑇𝑒𝑚𝑝𝑒𝑟𝑎𝑡𝑢𝑟𝑒 & Outlook = Rainfall) = 0.971 − 0.951 = 0.020

outlook

humidity

wind

YES

sunny

overcast

rainfall

NO

YES

Normal

High

YES

NO

2)

Nếu bị overfitting thì nên điều chỉnh max\_depth