ARTIFICIAL INTELLIGENCE

(WRITTEN PART OF ASSIGNMENT 5)

NAMES: FAIZAN SAFDAR ALI, AHSAN PARVEIZ

ROLL NO: 17100152, 17100078

1. Gini(True) = 1 – P(Yes)² - P(No)²

Gini (True) = 1 – (1/3)² - (2/3)²

Gini (True) = 1 – 1/9- 4/9

Gini(True) = 4/9

Gini (False) = 1 – P(Yes)² - P(No)²

Gini (False) = 1- (2/3)² - (1/3)²

Gini (False) = 1- 4/9 - 1/9

Gini (False) = 4/9

Gini [WINDY]= (1/3)\*4/9+(1/3)\*4/9

Gini [WINDY]= 4/9 = 0.444

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Play | No | | Yes | | No | | | No | | No | | Yes | | | | No | | Yes | | Yes |
| Humidity Sorted Values | 88 | | 88 | | 89 | | | 101 | | 101 | | 102 | | | | 103 | | 103 | | 103 |
| Humidity Split Positions | 87.5 | | | 88.5 | | | 95 | | | | 101.5 | | | 102.5 | | | 103.5 | | | |
| POSSIBILITIES | <= | > | | <= | | > | <= | | > | | <= | | > | <= | > | | <= | | > | |
| Yes | 0 | 4 | | 1 | | 3 | 1 | | 3 | | 1 | | 3 | 2 | 2 | | 4 | | 0 | |
| No | 0 | 5 | | 1 | | 4 | 2 | | 3 | | 4 | | 1 | 4 | 1 | | 5 | | 0 | |
| GINI | 0.494 | | | 0.492 | | | 0.481 | | | | 0.344 | | | 0.444 | | | 0.494 | | | |

1. The split position among ‘Windy’ and all splits of ‘Humidity’ that has the least GINI index is the binary split of ‘Humidity’ at 101.5. Therefore, this is the first best split.

>101.5

<=101.5

Humidity

1. Now computing the GINI index for the “Windy” attribute and all the splits of “Humidity” separately for both the sets.

|  |  |  |
| --- | --- | --- |
| CASE 1 (<=101.5) | | |
| Humidity | Windy | Play |
| 88 | True | No |
| 88 | False | Yes |
| 89 | True | No |
| 101 | True | No |
| 101 | True | No |

|  |  |  |
| --- | --- | --- |
| CASE 2 (>101.5) | | |
| Humidity | Windy | Play |
| 102 | False | Yes |
| 103 | False | No |
| 103 | True | Yes |
| 103 | True | Yes |

**CASE 1:**

GINI(True) = 1 – 1

GINI(True) = 0

GINI(False) = 1- 1

GINI(False) = 0

GINI[WINDY]= 0

Hence the best split can be splitting on the attribute ‘Windy’ for set 1.

NO

Yes

Not Windy

Windy

CASE 2 (>101.5)

Original

Set

Humidity <= 101.5

Humidity > 101.5

CASE 1(<=101.5) ((((<=101.5)

**CASE 2:**

GINI(True) = 1 – 1

GINI(True) = 0

GINI(False) = 1- (1/2)² - (1/2)²

GINI(False) = 1- 1/4 - 1/4

GINI(False) = 1/2

GINI[WINDY]= (1/2)\*1/2

GINI[WINDY]= 1/4 = 0.25

Now calculating GINI index for all possible binary splits of ‘Humidity’

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Play | Yes | | | No | | | Yes | | Yes |
| Humidity Sorted Values | 102 | | | 103 | | | 103 | | 103 |
| Humidity Split Positions | 101.5 | | 102.5 | | | 103.5 | | | |
| POSSIBILITIES | <= | > | <= | | > | <= | | > | |
| Yes | 0 | 3 | 1 | | 2 | 3 | | 0 | |
| No | 0 | 1 | 0 | | 1 | 1 | | 0 | |
| GINI | 0.375 | | 0.444 | | | 0.375 | | | |

Spliting on “Windy” is best because it has the least GINI index.

Original

Set

Humidity <= 101.5

Humidity > 101.5

**Yes**

Not Windy

Windy

Not Windy

Windy

CASE 2

**No**

**Yes**

CASE 1

‘Not Windy’ can further be split like:

Original

Set

Humidity <= 101.5

Humidity > 101.5

**Yes**

Not Windy

Windy

Not Windy

Windy

CASE 2

**No**

**Yes**

CASE 1

NOT WINDY SPLIT

**No**

**Yes**

Humidity > 102.5

Humidity <= 102.5

1. Entropy(parent) = -(4/9)log2(4/9) -(5/9)log2(5/9)

Entropy(parent) = 0.99

1. Entropy(True) = -(1/3)log2(1/3) -(2/3)log2(2/3)

Entropy(True) = 0.92

Entropy(False) = -(2/3)log2(2/3) -(1/3)log2(1/3)

Entropy(False) = 0.92

Entropy[windy] = (2/3)\*0.92+(1/3)\*0.92

Entropy[windy] = 0.92

Gain[windy]= Entropy(parent) – Entropy[windy]

Gain[windy]= = 0.07

1. Error(root) = 1- max(4/9, 5/9) = 1 – 5/9 = 4/9
2. Error(True) = 1 – 2/3 = 1/3

Error(False) = 1 – 2/3 = 1/3

Error[windy]= (6/9)\*1/3+(3/9)\*1/3 = 1/3 = 3/9

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Play | No | | Yes | | No | | | No | | No | | Yes | | | | No | | Yes | | Yes |
| Humidity Sorted Values | 88 | | 88 | | 89 | | | 101 | | 101 | | 102 | | | | 103 | | 103 | | 103 |
| Humidity Split Positions | 87.5 | | | 88.5 | | | 95 | | | | 101.5 | | | 102.5 | | | 103.5 | | | |
|  | <= | > | | <= | | > | <= | | > | | <= | | > | <= | > | | <= | | > | |
| Yes | 0 | 4 | | 1 | | 3 | 1 | | 3 | | 1 | | 3 | 2 | 2 | | 4 | | 0 | |
| No | 0 | 5 | | 1 | | 4 | 2 | | 3 | | 4 | | 1 | 4 | 1 | | 5 | | 0 | |
| Error | 4/9 | | | 4/9 | | | 4/9 | | | | 2/9 | | | 1/3 | | | 4/9 | | | |

1. The error answer of last part are subtracted from the root error but will not affect answer to this question as we have to select the lowest error rate. At 101.5, the error rate for ‘Humidity’ is least. Therefore, this is the best split.