NAME: **KUSHWANTH PASUMARTHI**

ROLL NO: **181CO238**

TOPIC: **DECISION TREE ALGORITHM**

**--Calculating Information Gain**

Gain (A) = Info (D) – $Info(D)_A$

Info (D) = -(3/10)log(3/10) – (7/10)log(7/10)

= -(0.3)(-1.736) – 0.7(-0.514)

= 0.881

| refund | Pi | Ni | I(pi, Ni) |
|---|---|---|---|
| yes | 0 | 3 | 0 |
| no | 3 | 4 | 0.985 |

Info (D) (refund) = (3/10)*I (0, 3) + (7/10)*I (3, 4)

= (3/10)*0 + (7/10)*0.984

= 0.689

| Marital status | Pi | Ni | I(Pi,Ni) |
|---|---|---|---|
| single | 2 | 2 | 1 |
| married | 0 | 4 | 0 |
| divorced | 1 | 1 | 1 |

Info (D) (marital status) = (4/10)*1 + (2/10)*1

= 0.4 + 0.2 = 0.6

| Taxable income | Pi | Ni | I(Pi, Ni) |
|---|---|---|---|
| >80k | 3 | 5 | 0.954 |
| <80k | 0 | 2 | 0 |

Info (D) (taxable income) = (8/10) I (3, 5) + (2/10)*I(0, 2)

= 0.7635

Gain (refund) = 0.881 – 0.689

= 0.191

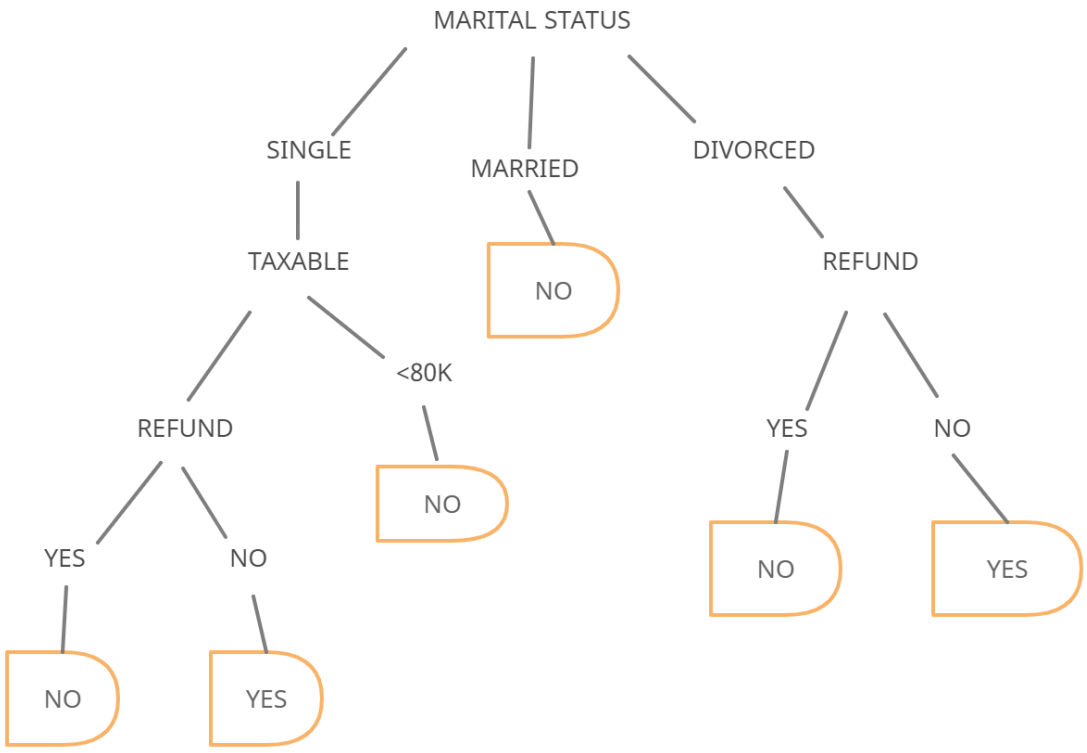
Gain (marital status) = 0.881 – 0.6

= 0.2813

Gain (taxable income) = 0.881 – 0.7635

= 0.117

**As per the gain value MARITAL STATUS is the root of decision tree**



**--calculating gain ratio:**

**Gain ratio (A) = Gain (A) / Split Info(A) ;**

Split Info (refund) = - (3/10)log(3/10) – (7/10)log(7/10)

= 0.881

Split Info (marital status) = - (4/10)log(4/10) – (4/10)log(4/10) – (2/10)log(2/10)

= 1.521

Split Info (taxable) = - (8/10)log(8/10) – (2/10)log(2/10)

= 0.721

Gain ratio (refund) = Gain (refund) / split info (refund)

= 0.191/0.881

= 0.216

Gain ratio (marital) = Gain (marital) / split info (marital)

= 0.281/1.521

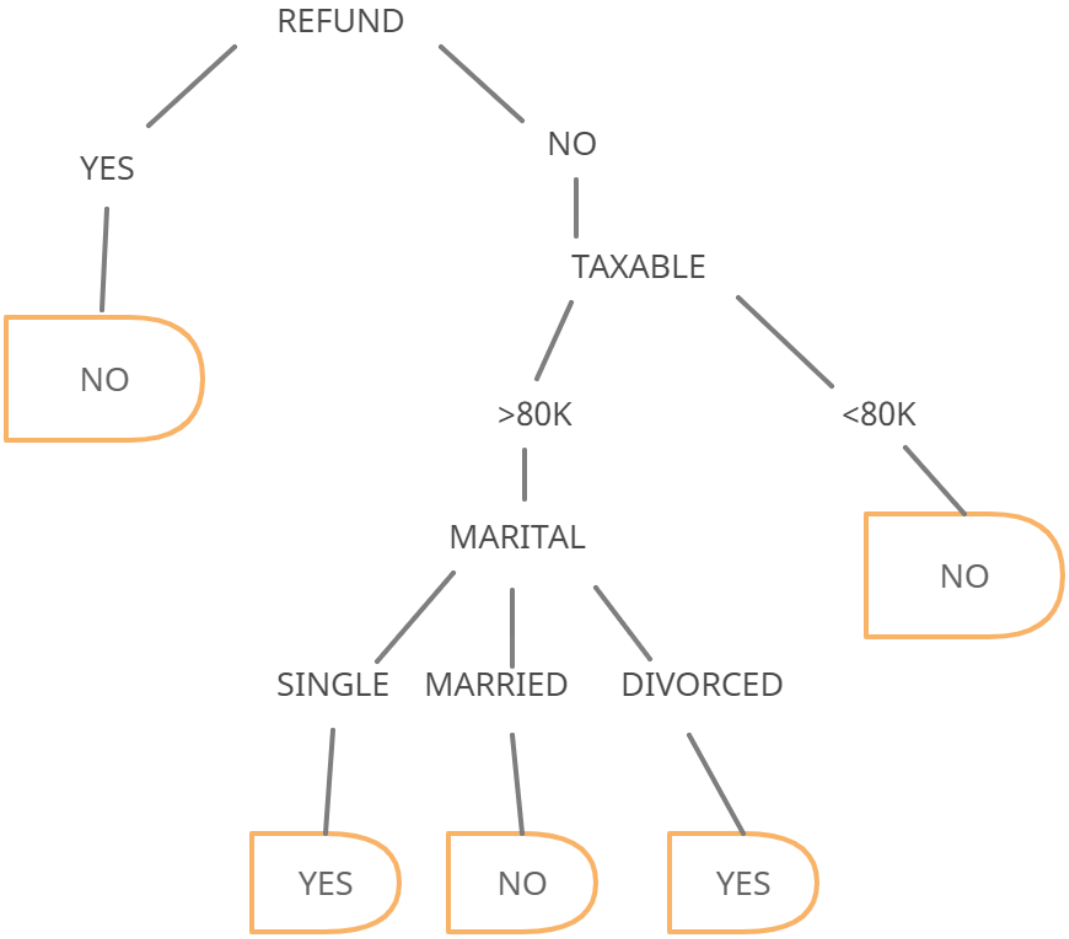
= 0.184

Gain ratio (taxable) = Gain (taxable) / split info (taxable)

= 0.117/0.721

= 0.162

**As per the gain ratio REFUND will be the root for decision tree**



**--Calculating Gini Index**

Gini (D) (refund) = 1 – (3/10)^2 – (7/10)^2

= 0.58

Gini (D) (marital status) { Splits into (single, married) and (divorced) }

= (8/10)*(1 – (4/10)^2 – (4/10)^2) + (2/10)*(1 – (1/10)^2 – (1/10)^2)

= 0.544 + 0.196

= 0.74

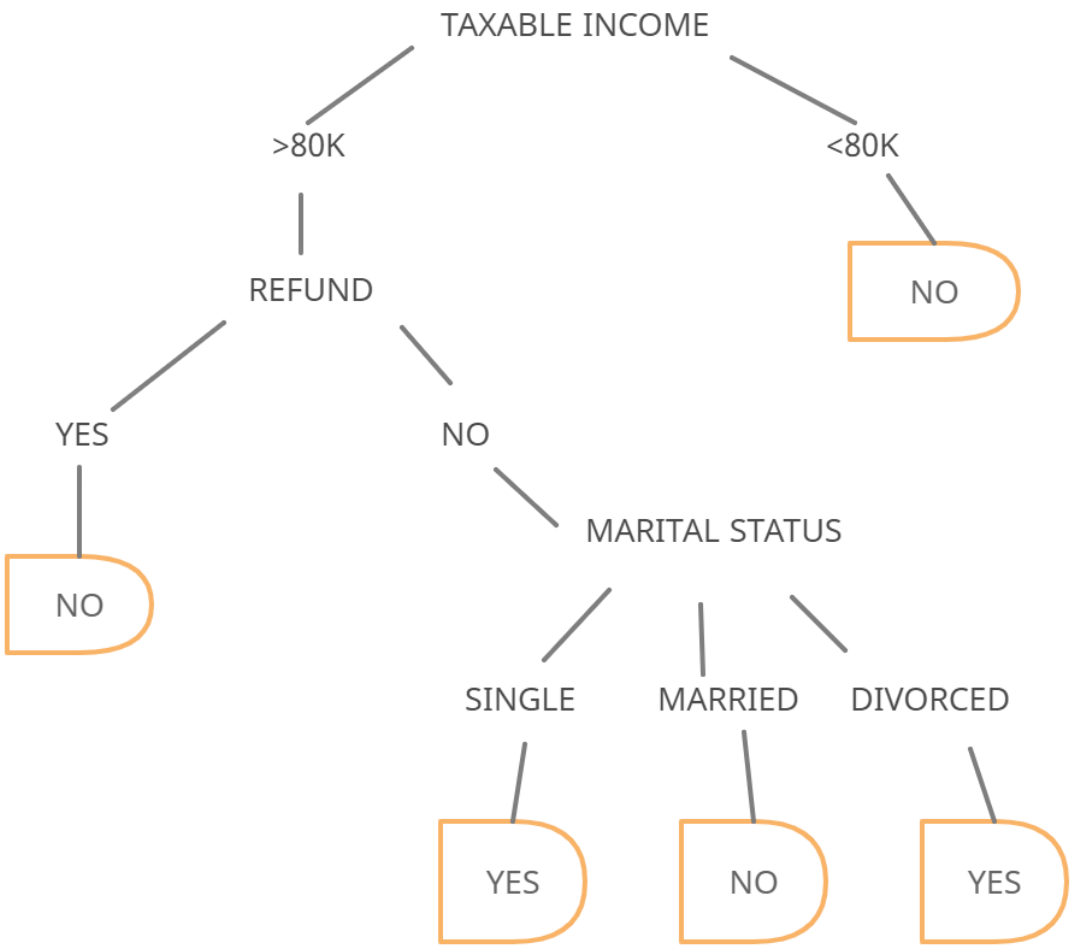
If we take (single) and (married, divorced) as partitions we are getting gini value as 0.848. So we should take the smallest split value.

Gini (D) (taxable income) = 1 – (2/10)^2 – (8/10)^2

= 0.32

**As per gini index values the lowest value term is the root i.e;**

**TAXABLE INCOME is the root for decision tree as per gini index**




➔ **All the logarithms are of base 2**