

# **Universidade São Judas Tadeu**

## **Inteligência artificial**

USJT\_2020/2 - CCP1AN-MCB3

Profª Gabriela Oliveira Biondi

### **Entrega 03**

Danilo Yuudi Hirata	819228171
Gabriel cardoso da silva	818136132
Henrique Olo Bardeja	81815270
Iuri Moura da Silva	818143167
Lucas Venceslau dos Santos Soares	818221111
Wagner Alves de Melo	818137692

## Resultados obtidos com a Árvore de Decisão:

Variável Target: **Native Country**

Condição: **Moradores dos Estados Unidos**

### Teste 1

train = :25000  
max\_depth = 7  
range: 1 a 19  
test = 25000:

#### Resultado:

Accuracy: 0.93 (+/- 0.01)  
Native Country = EUA: 27504 out of 30162 (91.19%)

Max depth: 1, Accuracy: 0.92 (+/- 0.01)  
Max depth: 2, Accuracy: 0.93 (+/- 0.01)  
Max depth: 3, Accuracy: 0.93 (+/- 0.01)  
Max depth: 4, Accuracy: 0.93 (+/- 0.01)  
Max depth: 5, Accuracy: 0.93 (+/- 0.00)  
Max depth: 6, Accuracy: 0.93 (+/- 0.01)  
Max depth: 7, Accuracy: 0.93 (+/- 0.01)  
Max depth: 8, Accuracy: 0.93 (+/- 0.01)  
Max depth: 9, Accuracy: 0.93 (+/- 0.01)  
Max depth: 10, Accuracy: 0.93 (+/- 0.01)  
Max depth: 11, Accuracy: 0.93 (+/- 0.01)  
Max depth: 12, Accuracy: 0.93 (+/- 0.00)  
Max depth: 13, Accuracy: 0.93 (+/- 0.00)  
Max depth: 14, Accuracy: 0.92 (+/- 0.00)  
Max depth: 15, Accuracy: 0.92 (+/- 0.01)  
Max depth: 16, Accuracy: 0.92 (+/- 0.01)  
Max depth: 17, Accuracy: 0.92 (+/- 0.01)  
Max depth: 18, Accuracy: 0.92 (+/- 0.00)  
Max depth: 19, Accuracy: 0.92 (+/- 0.01)

### Teste 2

train = :15000  
max\_depth = 5  
range: 1 a 9  
test = 15000:

Native Country = EUA: 27504 out of 30162 (91.19%)

### **Resultado:**

Accuracy: 0.93 (+/- 0.00)

Native Country = EUA: 27504 out of 30162 (91.19%)

Max depth: 1, Accuracy: 0.92 (+/- 0.00)

Max depth: 2, Accuracy: 0.93 (+/- 0.00)

Max depth: 3, Accuracy: 0.93 (+/- 0.00)

Max depth: 4, Accuracy: 0.93 (+/- 0.00)

Max depth: 5, Accuracy: 0.93 (+/- 0.00)

Max depth: 6, Accuracy: 0.93 (+/- 0.00)

Max depth: 7, Accuracy: 0.93 (+/- 0.00)

Max depth: 8, Accuracy: 0.93 (+/- 0.00)

Max depth: 9, Accuracy: 0.93 (+/- 0.00)

### **Teste 3**

train = :1000

max\_depth = 7

range: 1 a 14

test = 1000:

### **Resultado:**

Accuracy: 0.93 (+/- 0.00)

Native Country = EUA: 27504 out of 30162 (91.19%)

Max depth: 1, Accuracy: 0.92 (+/- 0.00)

Max depth: 2, Accuracy: 0.93 (+/- 0.00)

Max depth: 3, Accuracy: 0.93 (+/- 0.00)

Max depth: 4, Accuracy: 0.93 (+/- 0.00)

Max depth: 5, Accuracy: 0.93 (+/- 0.00)

Max depth: 6, Accuracy: 0.93 (+/- 0.00)

Max depth: 7, Accuracy: 0.93 (+/- 0.00)

Max depth: 8, Accuracy: 0.93 (+/- 0.00)

Max depth: 9, Accuracy: 0.93 (+/- 0.00)

Max depth: 10, Accuracy: 0.93 (+/- 0.00)

Max depth: 11, Accuracy: 0.93 (+/- 0.00)

Max depth: 12, Accuracy: 0.93 (+/- 0.00)

Max depth: 13, Accuracy: 0.93 (+/- 0.00)

Max depth: 14, Accuracy: 0.92 (+/- 0.00)

### **Teste 4**

train = :200

max\_depth = 7

range: 1 a 14

test = 200:

**Resultado:**

Accuracy: 0.93 (+/- 0.00)

Native Country = EUA: 27504 out of 30162 (91.19%)

Max depth: 1, Accuracy: 0.92 (+/- 0.00)

Max depth: 2, Accuracy: 0.93 (+/- 0.00)

Max depth: 3, Accuracy: 0.93 (+/- 0.01)

Max depth: 4, Accuracy: 0.93 (+/- 0.01)

**Teste 5**

train = :27000

max\_depth = 7

range: 1 a 9

test = 27000:

**Resultado:**

Accuracy: 0.93 (+/- 0.00)

Native Country = EUA: 27504 out of 30162 (91.19%)

Max depth: 1, Accuracy: 0.92 (+/- 0.01)

Max depth: 2, Accuracy: 0.93 (+/- 0.01)

Max depth: 3, Accuracy: 0.93 (+/- 0.01)

Max depth: 4, Accuracy: 0.93 (+/- 0.00)

Max depth: 5, Accuracy: 0.93 (+/- 0.00)

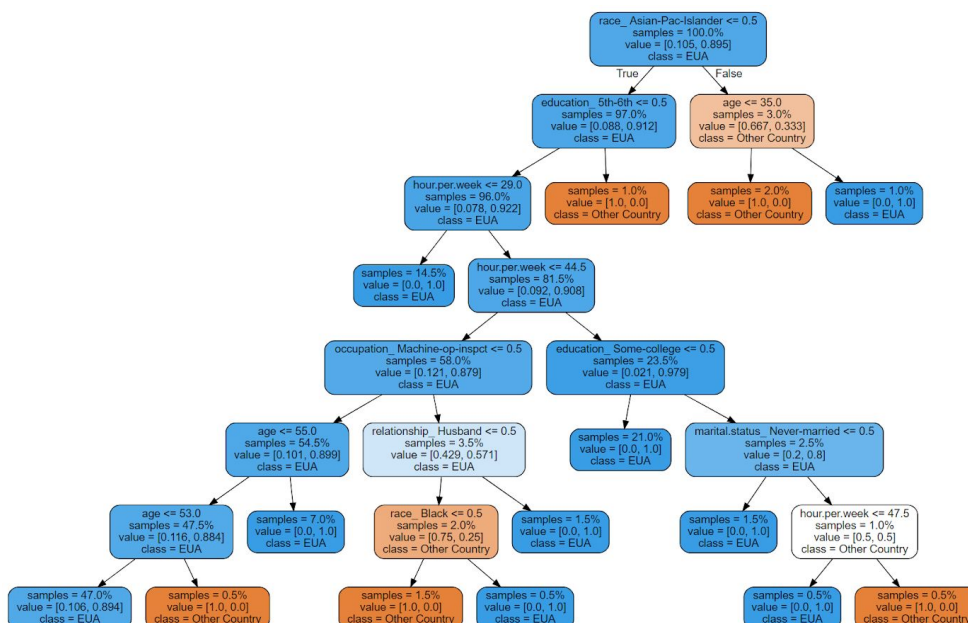
Max depth: 6, Accuracy: 0.93 (+/- 0.00)

Max depth: 7, Accuracy: 0.93 (+/- 0.00)

Max depth: 8, Accuracy: 0.93 (+/- 0.00)

Max depth: 9, Accuracy: 0.93 (+/- 0.00)

## Exemplo de teste da árvore do teste 4:



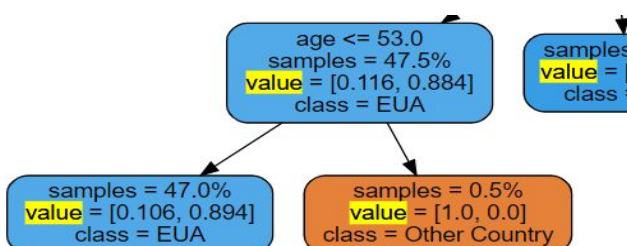
## Conclusão:

Nossa variável target é a native country, utilizamos os dados provenientes dela para analisar as variáveis restantes sobre as pessoas da base census e diferenciar informações dos moradores dos Estados Unidos das informações de moradores de outros países presentes na base.

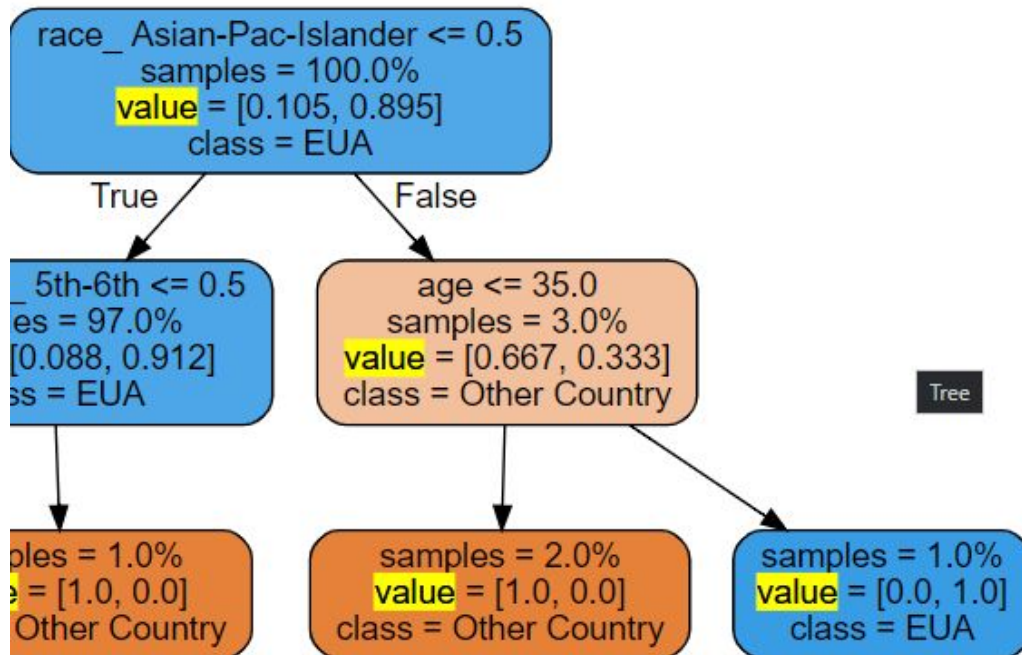
Realizamos 5 testes mudando o tamanho das amostras de treino, de teste e profundidade. Foi notada uma grande diferença entre o número de informações e tamanho da árvore conforme foram trocados os valores citados. De acordo com o teste, a acurácia varia entre 0.92 e 0.93, tendo como média 0.93. Também foi notado que o melhor intervalo de profundidade, ou seja, o que dispõe melhor acurácia é entre 2 e 13.

Com base na nossa variável target, no primeiro teste, foi possível notar que Estadunidenses formados no ensino superior e casados com a idade menor ou igual a 38 anos e meio tem salário maior do que 50 mil no ano.

## Exemplos:



A partir do teste 4, podemos concluir que pessoas com a idade menor ou igual a 53 somam 47% dos estadunidenses na base, e 0.5% de outros países.



Também foi possível concluir amostras de pessoas asiáticas da ilha do pacífico com a idade menor ou igual que 35 anos, são formadas 2% de moradores de outros países e 1% estadunidenses