

# Lista 1 IA

Daniel Salgado Magalhães - 821429

## Questão 1

Considerando-se a base de dados sobre “Esperar ou não pelo restaurante” (verificar base de dados ‘Restaurante.csv’ disponibilizada no CANVAS), pede-se:

1. Calcular o ganho de informação de cada atributo. Que atributo é a raiz da árvore?

**Resposta:** O atributo raiz são os clientes, que possuem o maior ganho de informação entre todos os atributos da tabela.

A	B	C	D	E	G	H	I	J	K	L	M	N	O	P	Q	R	S		
Exemplo	Altern	Bar	Sexo	fome	Clien	Preg	Chuva	Res	Tipo	Tempo	con			Total			Ganho		
x1	Sim	Não	Não	Sim	Alguns	RRR	Não	Sim	Francês	0-10	Sim	Preço RRRS	1/3	3	0,5283208	0,918295834	0,19570963		
x2	Sim	Não	Não	Sim	Cheio	R	Não	Não	Tailandês	30-60	Não	Preço RRRN	2/3	1/4	0,389975				
x3	Não	Sim	Não	Não	Alguns	R	Não	Não	Hamburger	0-10	Sim	Preço RRS	1	2	0	0			
x4	Sim	Não	Sim	Sim	Cheio	R	Sim	Não	Tailandês	out./30	Sim	Preço RRN	0	1/6	0				
X5	Sim	Não	Sim	Não	Cheio	RRR	Não	Sim	Francês	>60	Não	Preço RS	3/7	7	0,5238825	0,985228136			
X6	Não	Sim	Não	Sim	Alguns	RR	Sim	Sim	Italiano	0-10	Sim	Preço RN	4/7	3/5	0,4613457				
X7	Não	Sim	Não	Não	Nenhum	R	Sim	Não	Hamburger	0-10	Não			Total			Ganho		
X8	Não	Não	Não	Sim	Alguns	RR	Sim	Sim	Tailandês	0-10	Sim	Chuva SS	3/5	5	0,4421794	0,970950594	0,02072084		
X9	Não	Sim	Sim	Não	Cheio	R	Sim	Não	Hamburger	>60	Não	Chuva SN	2/5	3/7	0,5287712				
X10	Sim	Sim	Sim	Sim	Cheio	RRR	Não	Sim	Italiano	out./30	Não	Chuva NS	3/7	7	0,5238825	0,985228136			
X11	Não	Não	Não	Não	Nenhum	R	Não	Não	Tailandês	0-10	Não	Chuva NN	4/7	3/5	0,4613457				
X12	Sim	Sim	Sim	Sim	Cheio	R	Não	Não	Hamburger	30-60	Sim			Total			Ganho		
		Total			Ganho			Item	Tabela	Probabilidade	Entropia Pai	Ganho		Res SS	3/5	5	0,4421794	0,970950594	0,02072084
Alternativo SS	1/2	6	0,5	1	0			Decisão(Target)	Sim	1/2	0,5		Res SN	2/5	3/7	0,5287712			
Alternativo SN	1/2		0,5					Decisão(Target)	Não	1/2	0,5		Res NS	3/7	7	0,5238825	0,985228136		
Alternativo NS	1/2	6	0,5	1				Resultado			1	1	Res NN	4/7	3/5	0,4613457			
Alternativo NN	1/2		0,5											Total			Ganho		
		Total			Ganho								Tailandês S	1/2	4	0,5	1	0	
Bar SS	1/2	6	0,5	1	0								Tailandês N	1/2	1/3	0,5			
Bar SN	1/2		0,5										Italiano S	1/2	2	0,5	1		
Bar NS	1/2	6	0,5	1									Italiano N	1/2	1/6	0,5			
Bar NN	1/2		0,5										Francês S	1/2	2	0,5	1		
		Total			Ganho								Francês N	1/2	1/6	0,5			
Sex/Sab SS	2/5	5	0,5288	0,971	0,0207								Hamburguer S	1/2	4	0,5	1		
Sex/Sab SN	3/5	3/7	0,4422										Hamburguer N	1/2	1/3	0,5			
Sex/Sab NS	4/7	7	0,4613	0,9852										Total			Ganho		
Sex/Sab NN	3/7	3/5	0,5239										Tempo 0-10 S	2/3	6	0,389975	0,918295834	0,20751875	
		Total			Ganho								Tempo 0-10 N	1/3	1/2	0,5283208			
Fome SS	5/7	7	0,3467	0,8631	0,1957								Tempo 30-60 S	1/2	2	0,5	1		
Fome SN	2/7	3/5	0,5164										Tempo 30-60 N	1/2	1/6	0,5			
Fome NS	1/5	5	0,4644	0,7219									Tempo >60 S	0	2		0		
Fome NN	4/5	3/7	0,2575										Tempo >60 N	1	1/6	0			
		Total			Ganho								Tempo out./30 S	1/2	2	0,5	1		
Clientes AS	1	4	0	0	0,5409								Tempo out./30 N	1/2	1/6	0,5			
Clientes AN	0	1/3	0																
Clientes CS	1/3	6	0,5283	0,9183															
Clientes CN	2/3	1/2	0,39																
Clientes NS	0	2	0	0															
Clientes NN	1	1/6	0																

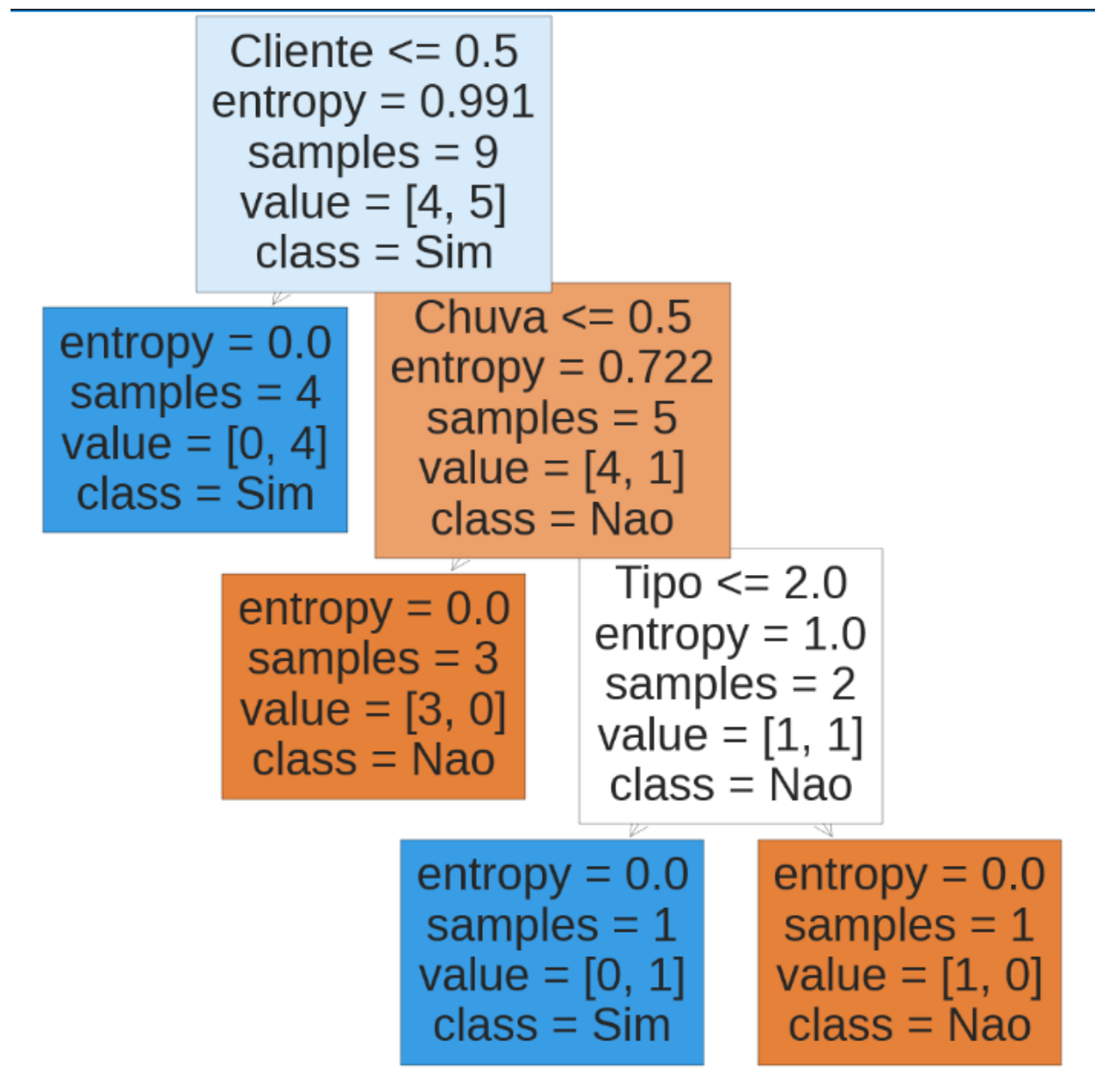
2. Que atributo estará no segundo nível da árvore. Faça os cálculos e apresente a árvore gerada até o segundo nível da árvore.

**Resposta:** O atributo que estará no segundo nível da árvore será o o atributo de tempo.

## Questão 2

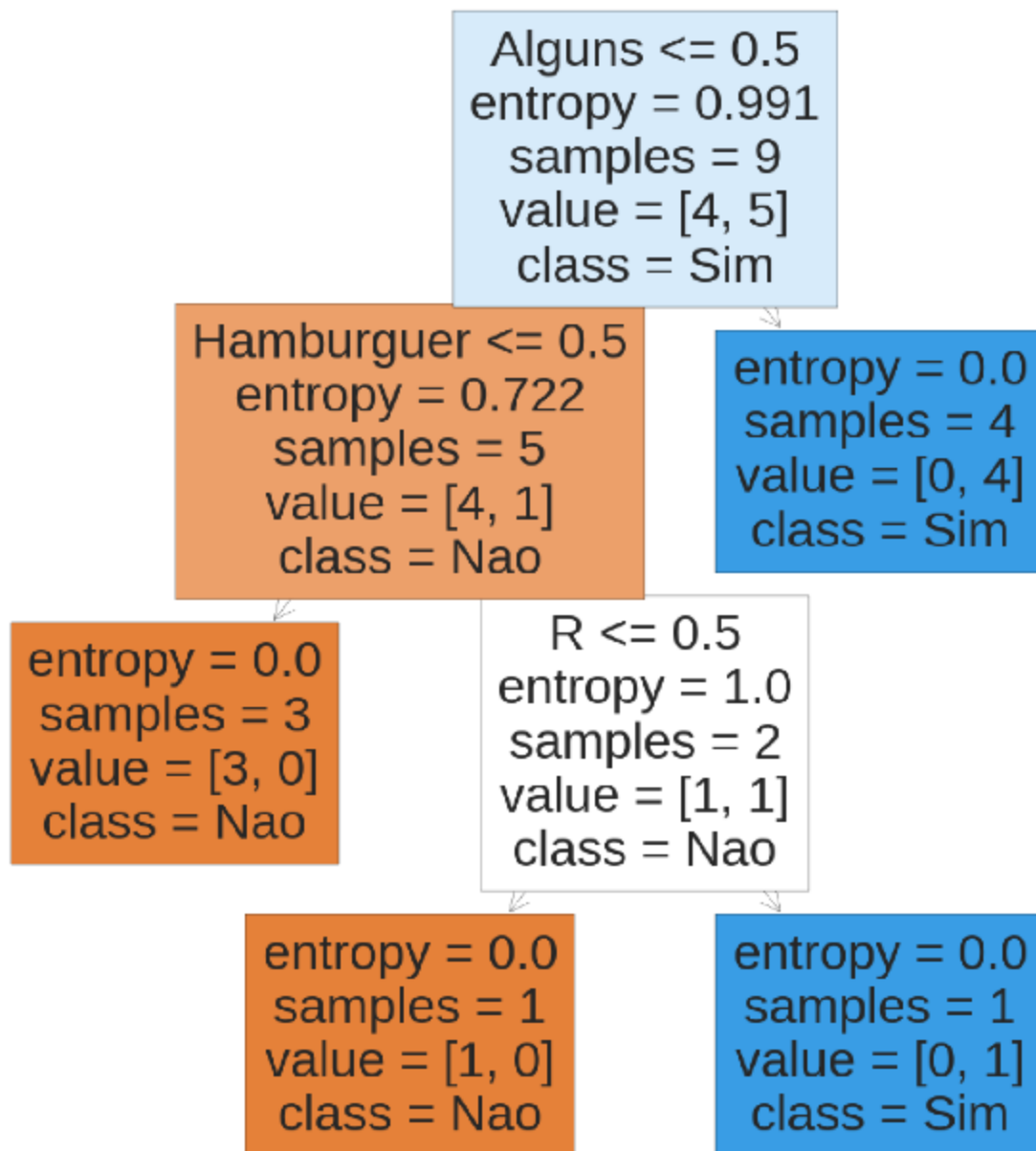
1. Com a codificação atual dos atributos de entrada, rode-os e plote a árvore de decisão obtida.

**Resposta:** Baixo plotado abaixo com o algoritmo de DecisionTree\_Restaurante



2. **Altere a codificação dos atributos (preço, cliente e tempo de espera) para um atributo nominal ordinal e faça os experimentos acima novamente. O que você observa? Houve ganho?**

**Resposta:** Alterando a codificação dos atributos de preços(R, RR e RRR), clientes(Alguns, Cheios e Nenhum) e tempo de espera(0-10, 30-60, out.30 e >60), o resultado não alterou drasticamente, devido a quantidade de objetos analisados, porém a estrutura da árvore em si alterou drasticamente, conforme indicado abaixo.



### Questão 3

#### 1. Quais as diferenças entre os algoritmos de árvore ID3 e C4.5?

**Resposta:** As diferenças entre os dois algoritmos são:

- ID3 trabalha apenas com atributos nominais, enquanto o C4.5 recebe dados numéricos , discretizando os mesmos;
- ID3 não recebe dados ausentes, enquanto o C4.5 é capaz de lidar com dados incompletos;
- ID3 não tem poda, enquanto o C4.5 possui. Neste caso, o C4.5 não permite que a árvore se ajuste demais aos dados de treinamento, reduzindo o tamanho da árvore removendo subárvores que adicionam pouco poder preditivo;
- ID3 tem dificuldade de lidar com dados nominais com muitas opções de resposta, enquanto o C4.5 utiliza a Razão de Ganho para escolher o melhor atributo que dividirá os dados em cada nó da árvore de decisão.

#### 2. Como o algoritmo C4.5 lida com os atributos de entrada que são numéricos?

**Resposta:** O algoritmo C4.5 lida com atributos numéricos ordenando as entradas, pegando os dois primeiros e fazendo a discretização deles, calculando a média entre estes dois números. Depois ele pega outro corte de números e vai fazendo esta operação até encontrar o corte ideal.