

Lista 1 - IA

Nome: Victor Ferraz de Moraes  
Matrícula: 802371

Exemplo	Alternativ o	Bar	Sex/Sa b	fome	Cliente	Preç o	Chuv a	Res	Tipo	Tempo	Vai esperar?
x1	Sim	Não	Não	Sim	Alguns	RRR	Não	Sim	Francês	0-10	Sim
x2	Sim	Não	Não	Sim	Cheio	R	Não	Não	Tailândes	30-60	Não
x3	Não	Sim	Não	Não	Alguns	R	Não	Não	Hamburger	0-10	Sim
x4	Sim	Não	Sim	Sim	Cheio	R	Sim	Não	Tailândes	10 - 30	Sim
x5	Sim	Não	Sim	Não	Cheio	RRR	Não	Sim	Francês	>60	Não
x6	Não	Sim	Não	Sim	Alguns	RR	Sim	Sim	Italiano	0-10	Sim
x7	Não	Sim	Não	Não	Nenhum	R	Sim	Não	Hamburger	0-10	Não
x8	Não	Não	Não	Sim	Alguns	RR	Sim	Sim	Tailândes	0-10	Sim
x9	Não	Sim	Sim	Não	Cheio	R	Sim	Não	Hamburger	>60	Não
x10	Sim	Sim	Sim	Sim	Cheio	RRR	Não	Sim	Italiano	10 - 30	Não
x11	Não	Não	Não	Não	Nenhum	R	Não	Não	Tailândes	0-10	Não
x12	Sim	Sim	Sim	Sim	Cheio	R	Não	Não	Hamburger	30-60	Sim

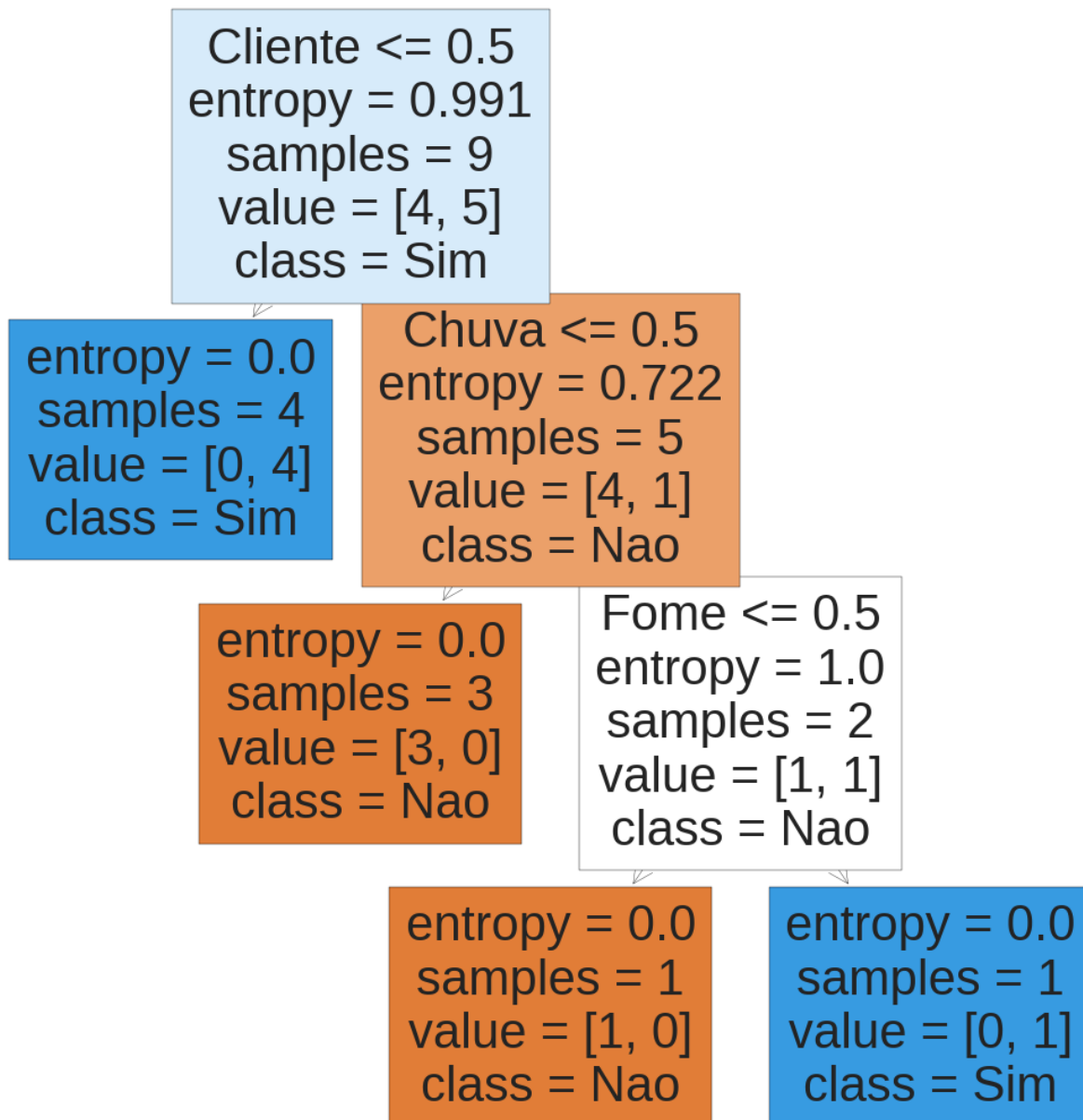
Questão 1)									
	Atributo (S)		Atributo (N)		Entropia Atributo	Ganho	Entropia Classe		
Alternativo	0.5		0.5		1	0	1		
Bar	0.5		0.5		1	0			
Sex/Sab	0.4045627477		0.5747164127		0.9792791604	0.02072083962			
fome	0.5034869983		0.3008033729		0.8042903712	0.1957096288			
Chuva	0.4045627477		0.5747164127		0.9792791604	0.02072083962			
Res	0.4045627477		0.5747164127		0.9792791604	0.02072083962			
	Nenhum	Alguns		Cheio					

Cliente	0	0	0.459147917	0.459147917	0.540852083				
	R	RR	RRR						
Preço	0.5747164127	0	0.2295739585	0.8042903712	0.1957096288				
	Francês	Tailândes	Hamburger	Italiano					
Tipo	0.1666666667	0.3333333333	0.3333333333	0.1666666667	1	0			
	0-10	10 - 30	30-60	>60					
Tempo	0.459147917	0.1666666667	0.1666666667	0	0.7924812504	0.2075187496			
Resposta Questão 1.1: O atributo de maior ganho é Cliente, logo ele será a Raiz a Árvore de Decisão.									
Questão 2)									
	Atributo (S)		Atributo (N)		Entropia Atributo	Ganho	Entropia Classe		
Alternativo	0.8091254954		0		0.8091254954	0.1091703387	0.9182958341		
Bar	0.459147917		0.459147917		0.9182958341	0			
Sex/Sab	0.8091254954		0		0.8091254954	0.1091703387			
fome	0.6666666667		0		0.6666666667	0.2516291674			
Chuva	0.3333333333		0.540852083		0.8741854163	0.04411041775			
Res	0		0.6666666667		0.6666666667	0.2516291674			
	R		RRR						
Preço	0.6666666667		0		0.6666666667	0.2516291674			
	Francês	Tailândes	Hamburger	Italiano					
Tipo	0	0.3333333333	0.3333333333	0	0.6666666667	0.2516291674			
	10 - 30		30-60	>60					

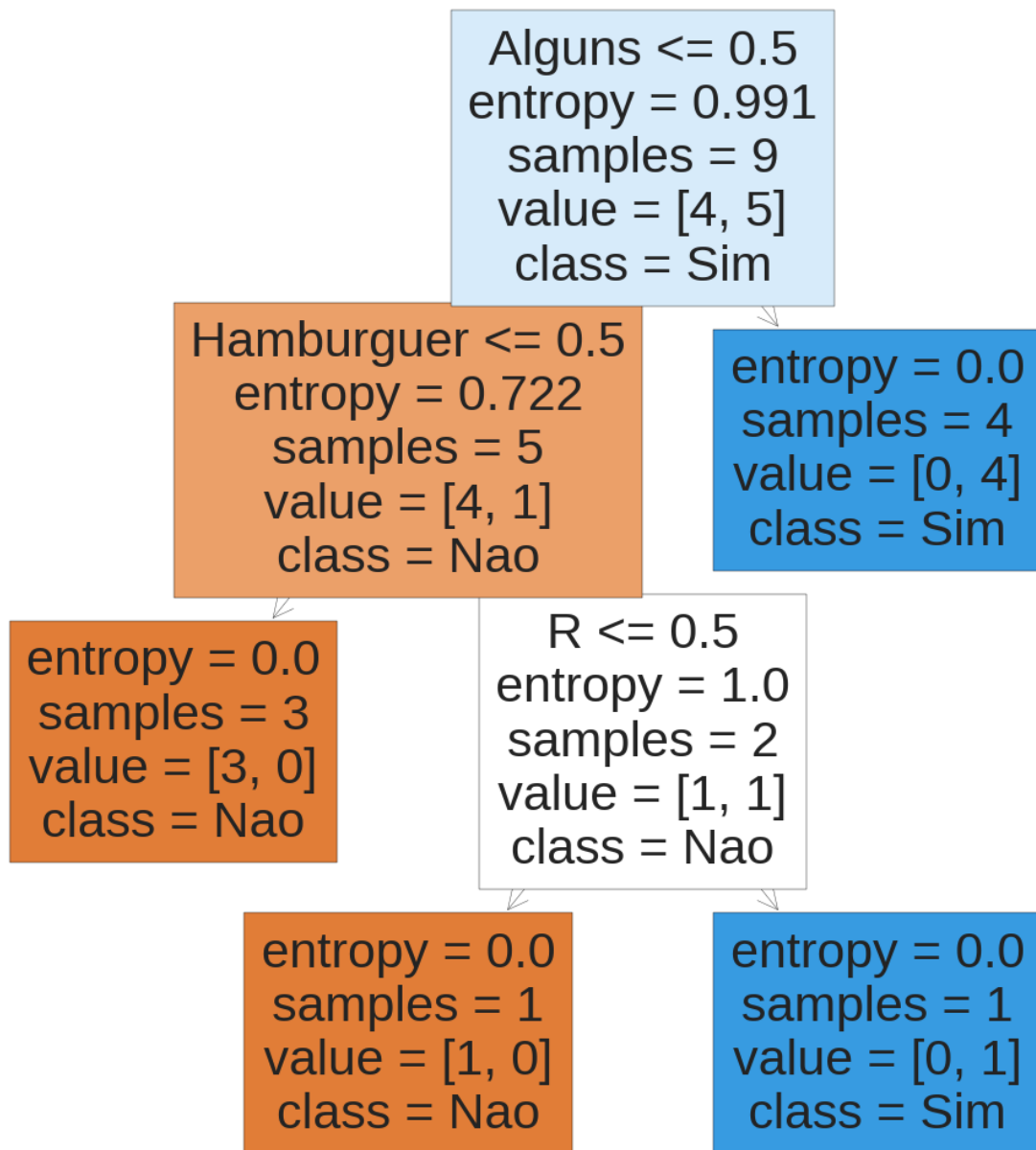
Tempo	0.3333333333	0.33333333 33	0	0.66666666 667	0.2516291674					
Resposta Questão 1.2: O atributo que estará no segundo nível da árvore poderá ser Res ou Fome.										

## Questão 2)

1)

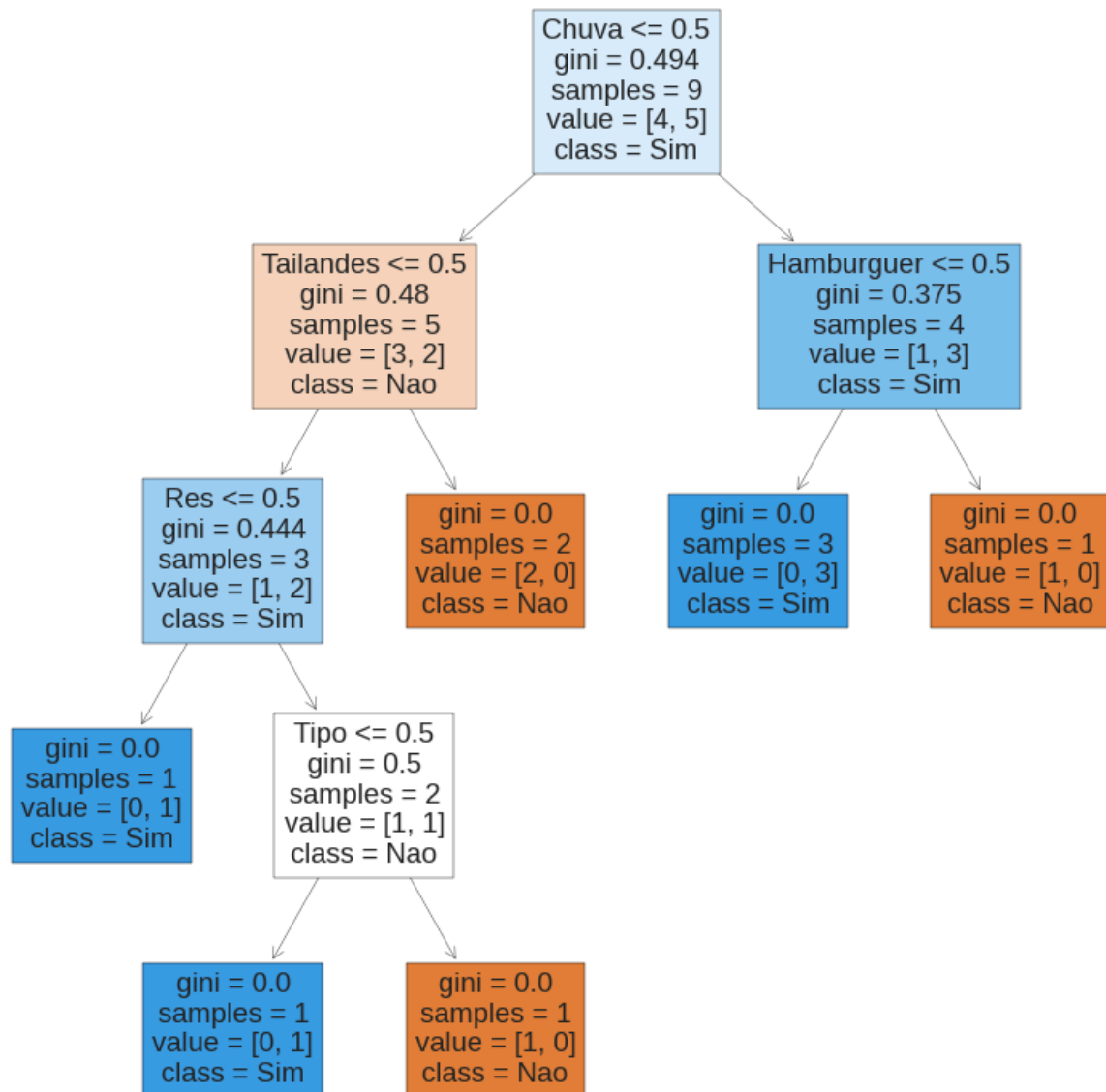


2)



**Resposta 2.2:** A primeira mudança notável é que a estrutura da Árvore de Decisão mudou drasticamente em comparação à primeira. Sobre o ganho, a raiz se tornou `Cliente(Alguns)`, porém não houve mudança em cima das métricas das previsões.

3)



**Resposta 2.3:** A árvore gerada apresentou uma mudança significativa em sua estrutura e ao ajustar os Hiperparâmetros, é possível perceber também que o melhor resultado foi gerado. O recall foi de 100% nas classes Sim e Não, além da alta precisão.

Nao	2	0
Sim	0	1
	Nao	Sim

### Questão 3)

#### Resposta 3.1:

**Tratamento de atributos contínuos :** o ID3 não é capaz de lidar diretamente com atributos contínuos. Por outro lado, o C4.5 pode dividir o domínio desses atributos em intervalos, permitindo maior flexibilidade na escolha de qual atributo escolher na divisão.

**Critério de divisão :** O ID3 é baseado apenas no ganho de informação no que se refere ao critério de seleção de atributos. Isso pode levar à escolha de um atributo com muitos valores, por exemplo. O C4.5 usa uma heurística chamada razão de ganho, pois normaliza o ganho de informação, evitando a seleção de atributos não informativos.

**Tratamento de Dados Ausentes:** O C4.5 tem um mecanismo embutido que lida eficientemente com dados ausentes. Mesmo se algumas informações estiverem faltando, a árvore pode ser construída. O ID3 não suporta essa função, tornando-o menos eficaz para conjuntos de dados complexos e incompletos.

**Poda de Árvores:** o C4.5 também executa técnicas de poda para reduzir o overfitting, o que melhora a generalização do modelo. Não existem mecanismos de poda análogos no ID3, e, portanto, as árvores geradas são menos generalizáveis e muito complexas.

**Eficiência e Tempo de Execução:** O C4.5 é mais eficiente e rápido em comparação com o ID3, sendo muito mais adequado para conjuntos de dados maiores.

**Resposta 3.2:** Por sua natureza, o algoritmo C4.5 é eficaz na manipulação de atributos de entrada numéricos e faz isso por meio de um processo chamado discretização, seguindo os seguintes passos:

Estabelece pontos de divisão dos valores desse atributo para criar intervalos de valores, em seguida o C4.5 analisa todos os valores numéricos possíveis para esse atributo e calcula a entropia ou o ganho de informação ao utilizar mais ou menos divisões. Após este processo, para cada ponto de divisão, calcula o ganho de informação, dividindo os dados em dois grupos (valores menores ou iguais ao ponto e aqueles com valores maiores). Assim que é determinado o ponto de divisão, o algoritmo cria uma regra de decisão que pode ser expressa como "se o atributo é menor ou igual ao valor X, então...". Portanto, permitindo, assim, o uso de árvores, mesmo que em qualquer nível a entrada correspondente ao atributo especificado para aquele ramo seja desconhecida.