CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO E AUTOMAÇÃO

LISTA DE EXERCÍCIOS 1

Morsinaldo de Azevedo Medeiros - 20220044322

Disciplina: DCA0200 - INTELIGÊNCIA ARTIFICIAL - T01 (2022.2 - 35M34)
Professor: ADRIÃO DUARTE DÓRIA NETO
Data: 20 de setembro de 2022

QUESTÃO 1: A tabela de dados abaixo ilustra a aplicação do método Naïve-Bayes. Um determinado banco deve decidir se um cliente deve ou não receber um empréstimo bancário em função da sua condição de bom ou mau pagador. Considerando os dados de treinamento abaixo, aplique o classificador Naive-Bayes, para atribuir a classe (rótulo) para os registros 12 e 13:

Registro	Casa Própria	Estado Civil	Possui Carro	Rendimentos	Bom Pagador
1	Sim	Solteiro	Sim	Alto	Não
2	Não	Casado	Sim	Médio	Não
3	Não	Solteiro	Não	Baixo	Não
4	Sim	Casado	Sim	Alto	Não
5	Não	Divorciado	Não	Médio	Sim
6	Não	Solteiro	Não	Baixo	Não
7	Sim	Casado	Sim	Alto	Sim
8	Não	Solteiro	Sim	Médio	Sim
9	Não	Casado	Sim	Baixo	Não
10	Não	Solteiro	Não	Médio	Sim
11	Sim	Divorciado	Não	Médio	Não
12	Não	Divorciado	Sim	Alto	?
13	Sim	Solteiro	Não	Médio	?

Em verde, estão destacados os dados de treinamento e em azul os dados de teste. Assim, o primeiro passo é calcular a tabela de probabilidades. Por questões de organização, utilizou-se as seguintes abreviações: CP = Casa Própria, EC = Estado Civil, PC = Possui Carro, R = Rendimentos e BP = Bom Pagador.

Tabela de probabilidades

- P(CP = Sim) = 4/11
- P(CP = Não) = 7/11
- P(EC= Solteiro) = 4/11
- P(EC = Casado) = 4/11
- P(EC = Divorciado) = 3/11
- P(EC = Divorciado) = 3/11
- P(PC = Não) = 5/11
- P(R = Alto) = 3/11
- P(R = Médio) = 5/11
- P(R = Baixo) = 3/11
- P(BP = Sim) = 4/11
- P(BP = Não) = 7/11
- P(CP =Sim | BP = Sim) = 1/4
- P(CP = Sim | BP = Não) = 3/7
- P(CP =Não | BP = Sim) = 3/4
- P(CP = Não | BP = Não) = 4/7
- P(PC =Sim | BP = Sim) = 2/4
- P(PC = Sim | BP = Não) = 4/7
- P(PC =Não | BP = Sim) = 2/4
- P(PC = Não | BP = Não) = 3/7
- P(EC = Solteiro | BP = Sim) = 2/4
- P(EC = Solteiro | BP = Não) = 2/7
- P(EC = Divorciado | BP = Sim) = 2/4
- P(EC = Divorciado | BP = Não) = 1/7
- P(R = Alto | BP = Sim) = 1/4
- P(R = Alto | BP = Não) = 2/7
- P(R = Médio | BP = Sim) = 1/4
- P(R = Médio | BP = Não) = 2/7

A seguir, precisa-se calcular a probabilidade dos registros 12 e 13 serem bons pagadores ou não a partir do teorema de Naive Bayes, o qual é descrito pela seguinte equação:

$$P(H|E) = \frac{P(E_1|H)\cdot P(E_2|H)\cdot \ldots \cdot P(E_n|H)\cdot P(H)}{P(E_1)\cdot P(E_2)\cdot \ldots \cdot P(E_n)} \quad (1)$$

Assim, para o registro 12, primeiro calculou-se a probabilidade de ser um bom pagador (BP), dado que ele não possui casa própria (CP), o estado civil (EC) é divorciado, possui carro (PC) e tem um rendimento (R) alto. Portanto:

$$P_{12S} = P(BP = Sim|CP = Não, EC = Divorciado, PC = Sim, R = Alto)$$

$$P_{12S} = \frac{P(CP=N|BP=S)\cdot P(EC=D|BP=S)\cdot P(PC=S|BP=S)\cdot P(R=A|BP=S)\cdot P(BP=S)}{P(CP=N)\cdot P(EC=D)\cdot P(PC=S)\cdot P(R=A)}$$

$$P_{12S} = \frac{3/4 \cdot 2/4 \cdot 2/4 \cdot 1/4 \cdot 4/11}{7/11 \cdot 3/11 \cdot 6/11 \cdot 3/11} = 0.6602 \simeq 66\%$$

A seguir, calculou-se a probabilidade de não ser um bom pagador dadas as mesmas características:

$$P_{12N} = P(BP = Não|CP = Não, EC = Divorciado, PC = Sim, R = Alto)$$

$$P_{12N} = \frac{P(CP=N|BP=N)\cdot P(EC=D|BP=N)\cdot P(PC=S|BP=N)\cdot P(R=A|BP=N)\cdot P(BP=N)}{P(CP=N)\cdot P(EC=D)\cdot P(PC=S)\cdot P(R=A)}$$

$$P_{12N} = \frac{4/7 \cdot 1/7 \cdot 4/7 \cdot 2/7 \cdot 7/11}{7/11 \cdot 3/11 \cdot 6/11 \cdot 3/11} = 0.3285 = 32.85\%$$

Como $P_{12S} > P_{12N}$, o cliente 12 é considerado um bom pagador pelo algoritmo Naive Bayes.

A seguir calculou-se a probabilidade de ser um bom pagador (BP) para o registro 13, dado que ele possui casa própria (CP), o estado civil (EC) é Solteiro, não possui carro (PC) e tem um rendimento (R) Médio. Portanto:

$$P_{13S} = P(BP = Sim|CP = Sim, EC = Solteiro, PC = Não, R = Médio)$$

$$P_{13S} = \frac{P(CP=S|BP=S)\cdot P(EC=Solteiro|BP=S)\cdot P(PC=N|BP=S)\cdot P(R=M|BP=S)\cdot P(BP=S)}{P(CP=S)\cdot P(EC=Solteiro)\cdot P(PC=N)\cdot P(R=M)}$$

$$P_{13S} = \frac{1/4\cdot 2/4\cdot 2/4\cdot 3/4\cdot 4/11}{4/11\cdot 4/11\cdot 5/11\cdot 5/11} = 0.6239 = 62.39\%$$

A seguir, calculou-se a probabilidade de não ser um bom pagador dadas as mesmas características:

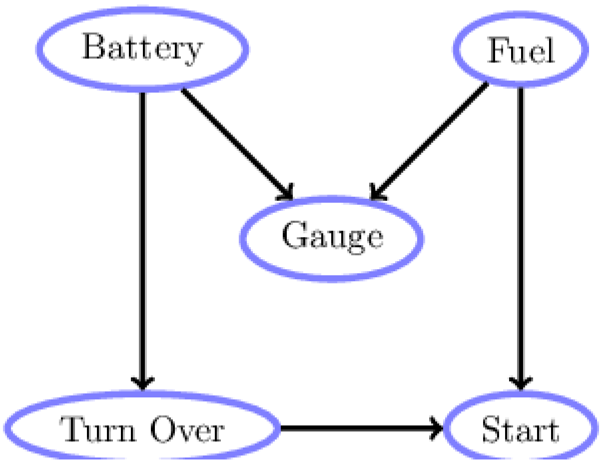
$$P_{13N} = P(BP = Sim|CP = Sim, EC = Solteiro, PC = Não, R = Médio)$$

$$P_{13N} = \frac{P(CP=S|BP=N)\cdot P(EC=Solteiro|BP=N)\cdot P(PC=N|BP=N)\cdot P(R=M|BP=N)\cdot P(BP=N)}{P(CP=S)\cdot P(EC=Solteiro)\cdot P(PC=N)\cdot P(R=M)}$$

$$P_{13N} = \frac{3/7\cdot 2/7\cdot 3/7\cdot 2/7\cdot 7/11}{4/11\cdot 4/11\cdot 5/11\cdot 5/11} = 0.3492 = 34.92\%$$

Como $P_{12S} > P_{12N}$, o cliente 13 também é considerado um bom pagador pelo algoritmo Naive Bayes.

QUESTÃO 2: A rede bayesiana abaixo concerne ao problema de partida de um carro, de uma forma bem simplificada (Extraído do Livro Bayesian Reasoning and Machine Learning - D. Barber).



Foram consideradas as seguintes abreviações: b = battery, g = gauge, f = fuel, t = turn over, s = start, fa = false, tr = true. As probabilidades referentes a rede bayesiana são dadas por:

p(b=bad)=0.05	p(f=empty)=0.1

p(g=empty|b=good,f=not empty)=0.05	p(g=empty|b=good,f=empty)=0.98
p(g=empty|b=bad,f=not empty)=0.07	p(g=empty|b=bad,f=empty)=0.97
p(t=fa|b=good)=0.1	p(t=fa|b=bad)=0.96
p(s=fa|t=tr, f=not empty)=0.01	p(s=fa|t=tr, f=empty)=0.92
p(s=fa|t=fa, f=not empty)=1.0	p(s=fa|t=fa, f=empty)=0.99

Um agente inteligente com base nas inferências, isto é, o cálculo da P(f=empty|s=no) (a probabilidade do tanque estar vazio dado que o carro não deu partida) e da P(b=bad|s=no) (a probabilidade da bateria estar descarregada e o carro não deu partida), deve decidir qual o problema mais provável pela não partida do carro. Apresente a solução e implemente os cálculos de forma computacional.

Analisando o gráfico da rede bayesiana, pode tirar as seguintes conclusões:

- Battery (b) não é filho de ninguém;
- Fuel (f) não é filho de ninguém
- Gauge (g) é filho de Battery (b) e Fuel (f)
- Turn Over (t) é filho de Battery (b)
- Start (s) é filho de Fuel (f)

De forma matemática, queremos calcular P(f=empty|s=fa). Assim, pela propriedade da independência condicional, que pela condição de Markov, podemos calcular a distribuição de probabilidade conjunta sobre todas as variáveis X_1, X_2, …, X_n da rede bayesiana através da seguinte fórmula:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(x_i | Pais(x_i)) \quad (2)$$

Assim, aplicando ao contexto da questão:

$$P(f = empty | s = fa) = \frac{P(f=empty, s=fa)}{P(s=fa)} = \frac{P(f=empty) \cdot P(s=fa | f=empty, t)}{P(s=fa)} \quad (2.1)$$

Calculando P(s=fa | empty, t):

$$P(s = fa | f = empty, t) = P(s = fa | f = empty, t = tr) \cdot P(t = tr | b) + P(s = fa | f = empty, t = fa) \cdot P(t = fa | b) \quad (2.2)$$

Mas não se conhece o valor de P(t = tr | b) e P(t = fa | b), assim

$$P(t = tr | b) = P(t = fa | b = bad) \cdot P(b = bad) + P(t = fa | b = good) \cdot P(b = good)$$

$P(t = tr|b) = 0.96 \cdot 0.05 + 0.1 \cdot (1 - 0.5) = 0.143$

Assim, $P(t = fa|b) = 1 - 0.143 = 0.857$

Substituindo os valores na equação 2.2,

$P(s = fa|f = empty, t) = 0.92 \cdot 0.857 + 0.99 \cdot 0.143 = 0.93001$

Da equação 2.1, não se conhece o valor de P(s=fa), o qual pode ser encontrado através da seguinte equação:

$$P(s = fa) = P(s = fa|\, f, t)$$

$$\begin{aligned} P(s = fa) = {} & P(s = fa|f = empty, t = fa) \cdot P(t = fa|b) \cdot P(f = empty) + \\ & P(s = fa|f = empty, t = tr) \cdot P(t = tr|b) \cdot P(f = empty) + \\ & P(s = fa|f = not\ empty, t = fa) \cdot P(t = fa|b) \cdot P(f = not\ empty) + \\ & P(s = fa|f = not\ empty, t = tr) \cdot P(t = tr|b) \cdot P(f = not\ empty) \end{aligned}$$

Substituindo os valores, temos:

$$\begin{aligned} P(s = fa) = {} & 0.99 \cdot 0.1 \cdot 0.143 + 0.92 \cdot 0.1 \cdot 0.857 + \\ & 1 \cdot 0.9 \cdot 0.143 + 0.01 \cdot 0.9 \cdot 0.857 \end{aligned}$$

$$P(s = fa) = 0.2294$$

Como todos os valores foram encontrados, pode-se calcular o resultado da equação 2.1

$$P(f = empty|s = fa) = \frac{P(f=empty) \cdot P(s=fa|f=empty,t)}{P(s=fa)} = \frac{0.1 \cdot 0.93001}{0.2294} = 0.4054$$

A seguir, irá se calcular a outra probabilidade requisitada pela questão:

$$P(b = bad|s = fa) = \frac{P(b=bad, s=fa)}{P(s=fa)} = \frac{P(b=bad) \cdot P(s=fa|\, f, t)}{P(s=fa)} \quad (2.3)$$

Para isso, precisa-se calcular o valor de $P(s = fa|\, f,\ t)$. Assim,

$$\begin{aligned} P(s = fa|\, f,\ t) = {} & P(s = fa|f = empty, t = tr) \cdot P(t = tr|b = bad) \cdot P(f = empty) + \\ & P(s = fa|f = empty, t = fa) \cdot P(t = fa|b = bad) \cdot P(f = empty) + \\ & P(s = fa|f = not\ empty, t = tr) \cdot P(t = tr|b = bad) \cdot P(f = not\ empty) + \\ & P(s = fa|f = not\ empty, t = fa) \cdot P(t = fa|b = bad) \cdot P(f = not\ empty) \end{aligned}$$

Como todas as probabilidades acima são conhecidas, pode-se substituir e encontrar o resultado.

$$P(s = fa|\, f,\ t) = 0.92 \cdot 0.1 \cdot (1 - 0.96) + 0.99 \cdot 0.1 \cdot 0.96 +$$

$$0.01 \cdot 0.9 \cdot (1 - 0.96) + 1 \cdot 0.9 \cdot 0.96$$

$$P(s = fa| f, t) = 0.96308$$

Substituindo os valores na equação 2.3,

$$P(b = bad|s = fa) = \frac{P(b=bad)\cdot P(s=fa| f, t)}{P(s=fa)} = \frac{0.05\cdot 0.96308}{0.2294} = 0.2099$$

QUESTÃO 3: Uma rede de crença (ou rede bayesiana), modela a relação entre as variáveis: oil (price of oil), inf (inflation) , eh (economy health), bp (British Petroleum Stock price), rt (retailer stock price). Cada variável tem dois estados (l:low) e (h:high), exceto a variável bp que tem adicionalmente o estado (n: normal). A rede de crença modela as variáveis de acordo com a tabela abaixo. (Extraído do Livro Bayesian Reasoning and Machine Learning - D. Barber)

eh(economy health)	
P(eh=low)	P(eh=high)
0.7	0.3

oil(price of oil)		
eh	oil=low	oil=high
low	0.9	0.1
high	0.05	0.95

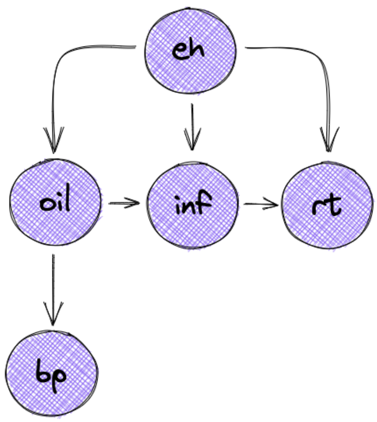
bp(British petroleum stock price)			
oil	bp=low	bp=high	bp=normal
low	0.9	0	0.1
high	0.1	0.5	0.4

rt(retailer stock price)			
inf	eh	rt=low	rt=high

low	low	0.9	0.1
low	high	0.1	0.9
high	low	0.1	0.9
high	high	0.01	0.99

inf(inflation)			
oil	eh	rt=low	rt=high
low	low	0.9	0.1
low	high	0.1	0.9
high	low	0.1	0.9
high	high	0.01	0.99

a) Determine o gráfico da rede de crença (rede bayesiana) para este problema.



b) Dado que a bp=n e rt=h, qual é a probabilidade de que a inflação seja alta?

Escrevendo de forma matemática,

$$P(inf = h|bp = n, rt = h) = \frac{P(inf=h, bp=n, rt=h)}{P(bp=n, rt=hi)} = \frac{P(inf=h|eh,oil)\cdot P(bp=n|oil)\cdot P(rt=h|eh,inf=h)}{P(bp=n, rt=hi)} \quad (3.1)$$

Calculando P(inf = h | eh, oil),

$$\begin{aligned} P(inf = h|eh, oil) = {} & P(inf = h|eh = l, oil = l) \cdot P(eh = l) \cdot P(oil = l|eh = l) + \\ & P(inf = h|eh = h, oil = h) \cdot P(eh = h) \cdot P(oil = h|eh = h) + \\ & P(inf = h|eh = l, oil = h) \cdot P(eh = l) \cdot P(oil = h|eh = l) + \end{aligned}$$

$$P(inf = h|eh = h, oil = l) \cdot P(eh = h) \cdot P(oil = l|eh = h)$$

$$P(inf = h|eh, oil) = 0.1 \cdot 0.7 \cdot 0.9 + 0.99 \cdot 0.3 \cdot 0.95 + 0.9 \cdot 0.7 \cdot 0.1 + 0.9 \cdot 0.3 \cdot 0.05$$

$$P(inf = h|eh, oil) = 0.42165$$

Calculando P(bp=n| oil),

$$P(bp = n|oil) = P(bp = n|oil = h) \cdot P(oil = h|eh) + P(bp = n|oil = l) \cdot P(bp = l|eh)$$

$$P(bp = n|oil) = P(bp = n|oil = h) \cdot [P(oil = h|eh = l) \cdot P(eh = l) + P(oil = h|eh = h) \cdot P(eh = h)] + P(bp = n|oil = l) \cdot [P(oil = l|eh = l) \cdot P(eh = l) + P(oil = l|eh = h) \cdot P(eh = h)]$$

$$P(bp = n|oil) = 0.4 \cdot [0.1 \cdot 0.7 + 0.95 \cdot 0.3] + 0.1 \cdot [0.9 \cdot 0.7 + 0.05 \cdot 0.3]$$
$$P(bp = n|oil) = 0.2065$$

Calculando P(rt=h | eh, inf=h) equação (3.2),

$$P(rt = h|eh, inf = h) = P(rt = h|inf = h, eh = l) \cdot P(inf = h|eh = l, oil) \cdot P(eh = l) + P(rt = h|inf = h, eh = h) \cdot P(inf = h|eh = h, oil) \cdot P(eh = h)$$

Para calcular o resultado da equação acima, precisa-se das probabilidades P(inf = h | eh = l, oil) e P(inf = h | eh = h, oil), as quais podem ser calculadas a partir das seguintes equações. Para a primeira probabilidade,

$$P(inf = h|eh = l, oil) = P(inf = h|eh = l, oil = l) \cdot P(oil = l|eh = l) \cdot P(eh = l) + P(inf = h|eh = l, oil = h) \cdot P(oil = h|eh = l) \cdot P(eh = l)$$

$$P(inf = h|eh = l, oil) = 0.1 \cdot 0.9 \cdot 0.7 + 0.9 \cdot 0.1 \cdot 0.7 = 0.126$$

Para a segunda,

$$P(inf = h|eh = h, oil) = P(inf = h|eh = h, oil = h) \cdot P(oil = h|eh = h) \cdot P(eh = h) + P(inf = h|eh = h, oil = l) \cdot P(oil = l|eh = h) \cdot P(eh = h)$$

$$P(inf = h|eh = l, oil) = 0.99 \cdot 0.95 \cdot 0.3 + 0.9 \cdot 0.05 \cdot 0.3 = 0.29565$$

Substituindo de volta na equação 3.2, temos que:

$$P(rt = h|eh, inf = h) = 0.9 \cdot 0.126 \cdot 0.7 + 0.99 \cdot 0.29565 \cdot 0.3 = 0.1672$$

Por fim, precisa-se calcular a probabilidade do denominador da equação 3, a qual está descrita a seguir:

$$P(bp = n,\ rt = hi) = P(bp = n|oil) \cdot P(rt = h|inf, eh)$$

Assim, calculando primeiramente P(bp = n | oil),

$$P(bp = n|oil) = P(bp = n|oil = l) \cdot P(oil = l|eh) + P(bp = n|oil = h) \cdot P(oil = h|eh)$$
$$P(bp = n|oil) = 0.2065$$

A seguir, calculando P(rt = h | inf, eh),

$$P(rt = h|inf, eh) = 0.1 \cdot 0.7 \cdot (1 - 0.126) + 0.9 \cdot 0.3 \cdot (1 - 0.29565) + 0.9 \cdot 0.7 \cdot 0.126 + 0.99 \cdot 0.3 \cdot 0.29565$$
$$P(rt = h|inf, eh) = 0.4185$$

Assim, substituindo todos os valores na equação 3,

$$P(inf = h|bp = n, rt = h) = \frac{0.42165 \cdot 0.2065 \cdot 0.1672}{0.2065 \cdot 0.4185} = 0.1685$$

QUESTÃO 4: Considere o problema de tomada de decisão caracterizado por uma sequência de eventos que podem ser apresentado por um grafo conhecido como rede de decisão. O problema em questão consiste das escolhas e das decisões por parte de uma empresa de petróleo. Uma determinada empresa petrolífera obteve a concessão para explorar uma certa região. Os estudos anteriores (testes preliminares) estimam a probabilidade de existir petróleo nessa região em 30 %. A companhia pode optar por um novo teste, que custa US$ 100.000,00, sendo que, se realmente existe petróleo, esse teste dirá com uma probabilidade de 0.85 que existe, e se realmente não existe, dirá com probabilidade 0.75 que não existe. Considerando que o custo de perfuração será de US$ 1.000.000,00 e que, se for encontrado petróleo, a companhia receberá US$ 20.000.000,00 pela produção. Considere, portanto os seguintes eventos e os seus complementos: (i) Evento T (a companhia faz o teste); (ii) Evento F (o teste é favorável à existência de petróleo; (iii) Evento P (a companhia perfura o poço); (iv) Evento E (existe petróleo).

a-) Construa a rede indicando os nós de decisões e os nós ao acaso (variáveis aleatórias). Considere as funções de utilidade, representadas por losangos, como sendo o lucro =receita - despesas, calculado em cada percurso da árvore.

b-) Determine em cada nó dos percursos da árvore a utilidade esperada.

c-)Usando o critério da utilidade máxima esperada, determine a melhor decisão.

d-) Qual o valor esperado do lucro da companhia se for tomada a melhor decisão? e-) Apresente também a solução deste problema através de um programa computacional e simule diferentes situações alterando o valor das probabilidades.

Observações:

(i) O evento inicial da árvore é se a companhia faz ou não faz o teste.

(ii) Para cada evento tem o seu complementar: Exemplo: T: Faz o teste, ¬T: Não faz o teste

(iii) Para o cálculo da utilidade esperada determine antes as probabilidades condicionais a posteriori com base no teorema de Bayes.

Resposta: O primeiro passo para a montagem da rede de decisão é realizar os cálculos das probabilidades de cada nó. Assim, determinando a probabilidade de haver ou não petróleo:

$$P(petróleo = Sim) = 0.3$$
$$P(petróleo = Não) = 0.7$$

Determinando a probabilidade de quando um novo é realizado.

$$P(nt = Sim \,|\, petróleo = Sim) = 0.85$$
$$P(nt = Não \,|\, petróleo = Não) = 0.75$$
$$P(nt = Não \,|\, petróleo = Sim) = 0.15$$
$$P(nt = Sim \,|\, petróleo = Não) = 0.25$$

Uma vez que um novo teste é realizado, é possível calcular a probabilidade do teste ser favorável ou não à existência de petróleo.

$$P(nt = Sim) = P(nt = Sim \,|\, petroleo = Sim) \cdot P(petroleo = Sim) + P(nt = Sim \,|\, petroleo = Não) \cdot P(petroleo = Não)$$

$$P(nt = Sim) = 0.85 \cdot 0.3 + 0.25 \cdot 0.7 = 0.43$$
$$P(nt = Não) = 1 - 0.43 = 0.57$$

Dado o resultado do novo teste, é possível determinar a probabilidade de encontrar ou não petróleo.

$$P(petróleo = Sim \,|\, nt = Sim) = \frac{P(petroleo=Sim, nt=sim)}{P(nt=Sim)} = \frac{P(nt=Sim \,|\, petroleo=Sim) \cdot P(petroleo=Sim)}{P(nt=Sim)}$$
$$P(petróleo = Sim \,|\, nt = Sim) = \frac{0.85 \cdot 0.3}{0.43} = 0.593$$

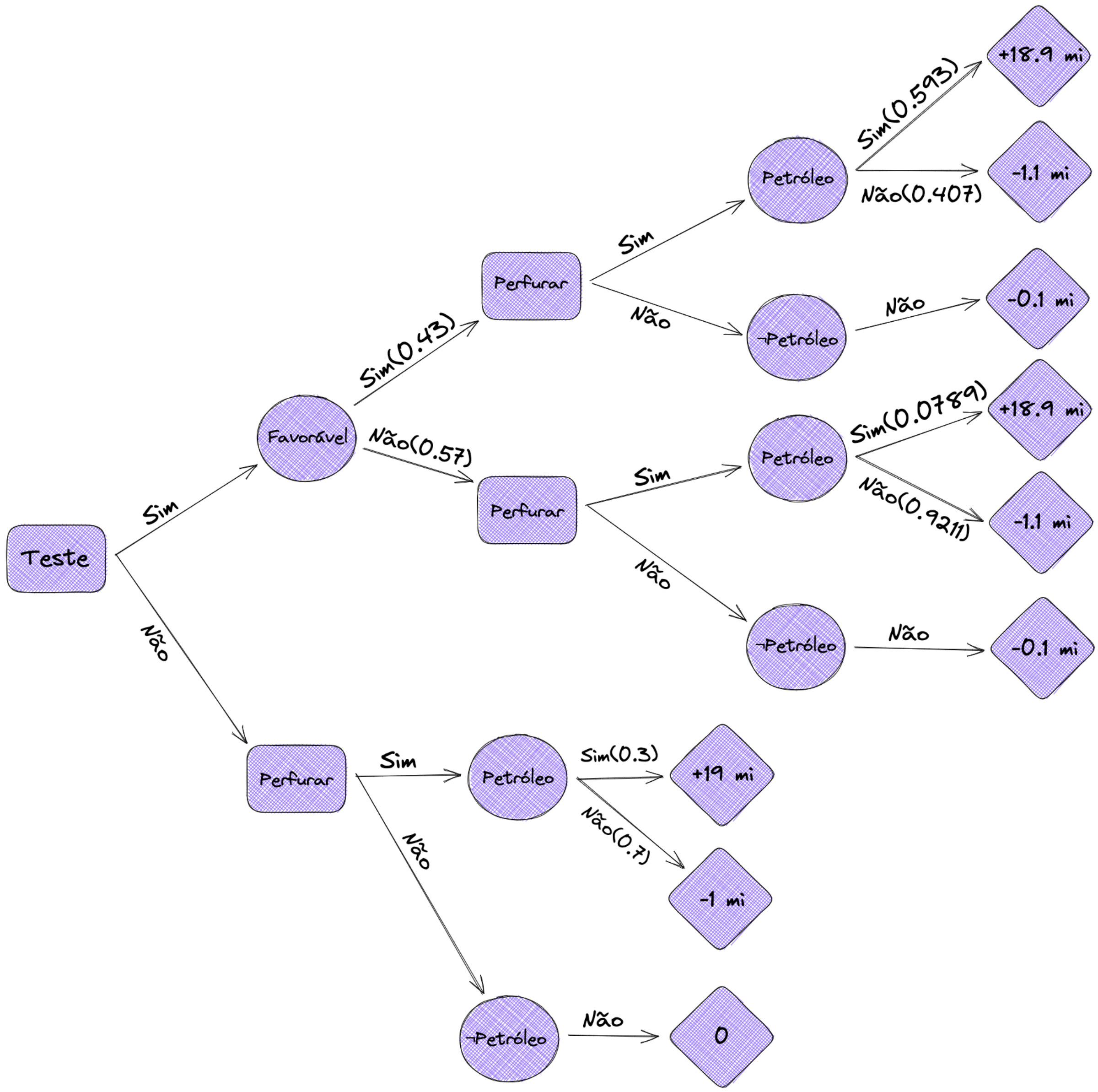
$$P(petróleo = Não \,|\, nt = Sim) = 1 - P(petróleo = Sim \,|\, nt = Sim) = 0.407$$

$$P(petróleo = Sim \,|\, nt = Não) = \frac{P(petroleo=Sim, nt=Não)}{P(nt=Não)} = \frac{P(nt=Não \,|\, petroleo=Sim) \cdot P(petroleo=Sim)}{P(nt=Não)}$$

$$P(petróleo\ =\ Sim\,|\,nt\ =\ Não)\ =\ \frac{0.15 \cdot 0.3}{0.57}\ =\ 0.0789$$

$$P(petróleo\ =\ Não\,|\,nt\ =\ Não)\ =\ P(petróleo\ =\ Sim\,|\,nt\ =\ Não)\ =\ 0.9211$$

Com as probabilidades calculadas, pode-se montar a árvore de decisão.



Por fim, precisa-se calcular a máxima utilidade esperada para encontrar a melhor decisão que a empresa pode tomar dadas as probabilidades. Para isso, é necessário calcular a utilidade de fazer ou não o teste.

$$U(\neg T)\ =\ 19.000.000\ \cdot\ 0.3\ -\ 1.000.000\ \cdot\ 0.7\ +\ 0\ =\ 5.000.000$$

$$U(T)\ =\ 0.43\ \cdot\ [0.593\ \cdot\ 18.900.000\ -\ 0.407\ \cdot\ 1.100.000\ -\ 100.000]\ +$$
$$0.57\ \cdot\ [0.0789\ \cdot\ 18.900.000\ -\ 0.9211\ \cdot\ 1.100.000\ -\ 100.000]$$

$U(T) = 4.799.260$

Assim, a melhor decisão é não fazer o teste, havendo uma recompensa ao final de 5.000.000.

QUESTÃO 5: Este exercício foi resolvido de forma computacional e pode ser encontrado no link: https://github.com/Morsinaldo/artificial_intelligence/blob/main/L1_Quest%C3%A3o5.ipynb

QUESTÃO 6: Apresente um estudo sobre as seguintes medidas de desempenho de classificadores de padrões: Acurácia, Precisão e F-Score.

RESPOSTA:

Acurácia: é uma métrica de avaliação de modelos de inteligência artificial em problemas de classificação. Ou seja, a acurácia é um valor numérico resultante da razão entre as previsões acertadas pelo modelo e o número total de previsão. Assim, podemos escrever a fórmula:

$$Acurácia = \frac{Número\ de\ previsões\ corretas}{Número\ total\ de\ previsões}$$

Para um problema de classificação binária, na qual temos duas classes, podemos calcular a acurácia da seguinte forma:

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN},$$

onde VP = Verdadeiro positivo, VN = Verdadeiro negativo, FP = Falso positivo e FN = Falso negativo. Apesar da demonstração ser para um problema de classificação de duas classes, esta métrica pode ser utilizada para problemas de classificação multiclasses.

Precisão: é uma métrica calculada a partir da razão entre a quantidade de verdadeiros positivos sobre o número total de classificações positivas, isto é, a soma de verdadeiros positivos e falso positivos. Assim, quanto mais previsões verdadeiras corretas, mais a precisão do modelo se aproxima de 1, ao passo que se aproxima de 0 à medida que o número de falso positivos tende a aumentar. Assim, podemos escrever a equação da precisão da seguinte maneira:

$$Precisão = \frac{VP}{VP + FP}$$

Uma situação em que essa métrica é de suma importância é na realização de empréstimos dado o histórico de um mutuário, pois caso o modelo classifique a pessoa como sendo boa pagadora, sendo que na verdade ela é mau pagadora (falso positivo), o modelo irá fazer com que os investidores percam dinheiro.

F-Score: É uma métrica utilizada para a avaliação de modelos de classificação e é definida como a média harmônica entre a precisão e o recall. Também conhecido como sensibilidade, o recall é outra métrica de avaliação de modelos que é definida pela razão entre os verdadeiros positivos e a soma dos verdadeiros positivos e falso negativos. As fórmulas do recall e do F-Score são definidas pela equação abaixo:

$$Recall = \frac{VP}{VP + FN}$$

$$F_{\beta} = (1 + \beta^2).\frac{Precisão.Recall}{(\beta^2.precisão) + recall}$$

Os mais comuns são: o F1-Score fornece pesos iguais para o precisão e para o recall, o F2-Score que fornece um peso maior para o recall e o F0.5-Score que fornece um peso maior para a precisão. Em relação às suas aplicações, o F-Score é comumente utilizado em sistemas que utilizam mecanismos de busca e processamento de linguagem natural.

TRABALHO 1: Pesquise e apresente um trabalho sobre o algoritmo Naïve-Bayes para a detecção de Spam em mensagens de email ou para classificar páginas de texto como interessante ou não interessante com base em um tema de interesse (esporte, política, ...etc.) presente nas palavras que aparecem nas páginas.

O notebook com o desenvolvimento do trabalho pode ser encontrado no link: https://github.com/Morsinaldo/artificial_intelligence/blob/main/Trabalho_01_IA_Naive_Bayes.ipynb

TRABALHO 2: Pesquise e apresente um trabalho sobre Random Forest com uma aplicação de livre escolha.

O notebook com o desenvolvimento do trabalho pode ser encontrado no link: https://github.com/Morsinaldo/artificial_intelligence/blob/main/Trabalho_02_IA_Random_Forest.ipynb