Nome: ALVARO MAIA CHAVES
Pontuação Total da Avaliação: 6.50 pontos
==========
Correção da Questão 1:
Pergunta: 1a) Uma empresa de recrutamento on-line deseja avaliar o potencial de seus assinantes que estão em busca de emprego. Para isso, reuniu um conjunto de dados contendo informações sobre candidatos que aplicaram para uma vaga, juntamente com o resultado final da análise da empresa empregadora: "Aceito" ou "Rejeitado". Cada candidato possui três atributos: Experiência (Alta/Baixa); Graduação (Sim/Não); Habilidades Técnicas (Boa/Ruim). A empresa deseja construir uma árvore de decisão para prever se um candidato será aceito ou rejeitado com base nesses atributos. Explique como a entropia e o ganho de informação são utilizados para decidir qual atributo deve ser escolhido como raiz da árvore de decisão.
==========Rubrica(s)
('Citar e/ou explicitar o cálculo da entropia inicial de "Aceito" ou "Rejeitado" antes de realizar a partição.', 0, 8)
('Citar e/ou explicar o calculo de Information Gain (IG) como sendo IG = Entropia(Antes da partição) - Entropia(Após partição), considerando que o maior IG define qual atributo será escolhido', 0.8)
('Citar e/ou explicitar o cálculo da entropia final de "Aceito" ou "Rejeitado" considerando as partições possíveis no atributo que está sendo analisado.', 0, 8)
==========Resposta do estudante da questão 1:
1a) A entropia serve para saber se um conjunto de dados sera facilmente classificavel ou seja uma entropia alta significa que o conjunto de dados Ã© dificil classificar os dados pois hÃ¡ muitas classes diferentes, ja uma entropia baixa significa uma facilidade ao classificar ja que representa uma quantidade de dados concentrada em poucas classes.
Assim utilizando a formula da entropia H(a)=-P(a1)*log2(P(a2))-P(a1)*log2(P(a2)) possibilita calcular a entropia inicial do conjunto e logo depois a entropia de cada atributo. Apos isso Ã© calculado a entropia media final com os resultados dos calculos anteriores para enfim poder calcular o ganho de informaÃ§Ã£o com a entropia inicial ou seja Ganho=H(inicial)-h(media final), mostrando assim qual atributo Ã© a melhor escolha para aquela raiz.
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante apresenta uma compreensão básica do conceito de entropia e seu papel na construção de uma árvore de decisão, mas carece de precisão e detalhamento em relação às rubricas fornecidas. O estudante menciona a fórmula da entropia, mas não a aplica corretamente ao contexto da questão, especialmente em relação ao cálculo da entropia inicial e final. Além disso, a explicação do ganho de informação é parcial, pois não é claro como o estudante pretende aplicar essa noção para decidir qual atributo deve ser escolhido como raiz da árvore de decisão.
==========Feedback:
Para melhorar a resposta, o estudante deve começar citando e explicitando o cálculo da entropia inicial de "Aceito" ou "Rejeitado" antes de realizar a partição, considerando a probabilidade de cada classe. Em seguida, deve explicar como o cálculo de Information Gain (IG) é feito, destacando que IG = Entropia(Antes da partição) - Entropia(Após partição), e que o maior IG define qual atributo será escolhido. Além disso, o estudante deve calcular a entropia final de "Aceito" ou "Rejeitado" considerando as partições possíveis no atributo que está sendo analisado, apresentando os passos de cálculo de forma clara e concisa. Isso permitirá uma compreensão mais profunda do processo de construção da árvore de decisão.
==========Pontuação:
1.2
==========
Correção da Questão 2:
Pergunta: 1b) Uma empresa de recrutamento on-line deseja avaliar o potencial de seus assinantes que estão em busca de emprego. Para isso, reuniu um conjunto de dados contendo informações sobre candidatos que aplicaram para uma vaga, juntamente com o resultado final da análise da empresa empregadora: "Aceito" ou "Rejeitado". Cada candidato possui três atributos: Experiência

(Alta/Baixa); Graduação (Sim/Não); Habilidades Técnicas (Boa/Ruim). A empresa deseja construir uma árvore de decisão para prever se um candidato será aceito ou rejeitado com base nesses atributos. Suponha que a entropia inicial do conjunto seja 0.94. Após dividir os dados com base no atributo Experiência, obtemos: Candidatos com Experiência: 42 Aceitos e 7 Rejeitados; Candidatos sem Experiência: 12 Aceitos e 78 Rejeitados. Calcule o ganho de informação desse atributo e interprete o resultado.
==========Rubrica(s)
('Citar e/ou explicitar o cálculo da entropia do grupo de amostras formadas para H(Experiência == "Alta")=-(42/49)*log2(42/49)-(7/49)*log2(7/49)~0.59', 1.0)
('Citar e/ou explicitar o cálculo da entropia do grupo de amostras formadas para H(Experiência != "Alta")=-(12/90)*log2(12/90)-(78/90)*log2(78/90)~0.57', 0.8)
('Citar e/ou explicar que um ganho de informação maior que zero demonstra maior homogeneidade dos subconjuntos após a partição "Experiencia Alta".', 0.5)
('Citar e/ou explicar o cálculo do ganho de informação (Information Gain, IG) como sendo IG("Experiencia Alta")~0.94-0.58=0.36', 1.0)
('Citar e/ou explicar o cálculo da entropia ponderada após a partição Experiência == "Alta", H("Experiencia Alta")=(49/139)*0.59+(90/139)*0.57~0.58', 0.8)
==========Resposta do estudante da questão 2:
1b) Como ja foi dada a entropia inicial, nÃ£o hÃ¡ a necessidade de calculÃ¡-la
H(inicial)=0,94
Considerando: Com Experiencia = E, Sem experiencia = SE, Aceito = A e Rejeitado= R
H(E|A) = 42
H(E|R) = 7
H(E|A) = - P(E|A)*log2(P(E|A)) - P(E|R)*log2(P(E|R))
H(E|A) = -(42/49)*log2(42/49) -(7/49)*log2(7/49)
H(E|A) = -0,86*-0,22 -0,14*-2,80
H(E|A) = 0,19 + 0,392
H(E|A) = 0,582
H(SE|A) = 12
H(SE|R) = 78
H(SE|A) = - P(SE|A)*log2(P(SE|A)) - P(SE|R)*log2(P(SE|R))
H(SE|A) = -(12/90)*log2(12/90) -(78/90)*log2(78/90)
H(SE|A) = -0,13*-2,90 -0,86*-0,206
H(SE|A) = 0,337 + 0,177
H(SE|A) = 0,554
Agora a entropia media final
Ja que 42+7+12+78 = 139 Ã© o numero total de canditados nesse conjunto, entÃ£o
H(E) = (49/139)*0,582 + (90/139)*0,554
H(E) = 0,564
Para calcular agora o ganho de informaÃ§Ã£o temos,
H(A) = H(inicial) - h(media final)
H(A) = 0,94 - 0,564
H(A) = 0,376
Assim temos um ganho de informaÃ§Ã£o de 0,376 oque mostra que com o atributo experiencia existe uma certa influencia de 0,376 na hora de se definir um cadidato como aceito ou rejeitado.
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante apresenta cálculos corretos para a entropia inicial, entropia dos subconjuntos e ganho de informação. No entanto, há pequenos erros de arredondamento nos cálculos da entropia dos subconjuntos. O estudante calculou corretamente a entropia ponderada após a partição Experiência == "Alta" e também obteve o ganho de informação. A resposta demonstra conhecimento sobre o assunto, mas poderia ser mais precisa nos cálculos e explicação dos conceitos.
==========Feedback:
O estudante demonstrou conhecimento básico, mas precisa de melhoria nos cálculos exatos e na apresentação. Além disso, seria útil incluir uma explicação mais clara sobre o que o ganho de

informação representa e como ele é interpretado no contexto do problema. Também, seria interessante se o estudante discutisse o impacto da Experience Alta na decisão de aceitar ou rejeitar os candidatos.
==========Pontuação:
2.3
==========
Correção da Questão 3:
Pergunta: 2) Uma empresa de e-commerce deseja prever se um cliente comprará ou não um produto após visualizar a página do item. Para isso, foi analisado um conjunto de 200 interações de clientes e coletados os seguintes atributos: Tempo na Página (Curto ou Longo); Dispositivo (Mobile ou Desktop); Origem do Tráfego (Orgânico ou Pago). A tabela a seguir resume os dados coletados: ||Característica | Comprou (Sim) | Não Comprou (Não) || Tempo na Página = Longo | 60 | 30 || Tempo na Página = Curto | 20 | 90 || Dispositivo = Desktop | 50 | 50 || Dispositivo = Mobile | 30 | 70 || Origem do Tráfego = Orgânico | 40 | 40 || Origem do Tráfego = Pago | 40 | 60||. Sabemos que 80 clientes compraram o produto e 120 não compraram. Suponha que um novo usuário acessa a página do produto com as seguintes características: Tempo na Página = Longo; Dispositivo = Desktop; Origem do Tráfego = Orgânico. Considere: P(A|B,C,...,Z)=(P(A)P(B|A)P(C|A)...P(Z|A))/(P(B)P(C)...P(Z)).
==========Rubrica(s)
('Citar e/ou explicitar o cálculo P(Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico)=[(80/200)*(60/80)*(50/80)*(40/80)]/[(90/200)*(100/200)*(80/180)]~0.94 ou memso a simplificação P(Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico)=[(80/200)*(60/80)*(50/80)*(40/80)]~0.094, desconsiderando o denominador P(Tempo Longo)*P(Dispositivo Desktop)*P(Tráfego Orgânico) tendo em vista que irá comparar com a P(Não Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico) com a mesma simplificação.', 2.0)
('Citar e/ou explicitar o cálculo P(Não Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico)=[(120/200)*(30/120)*(50/120)*(40/100)]/[(90/200)*(100/200)*(80/180)]=0.25 ou mesmo a simplificação P(Não Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico)=[(120/200)*(30/120)*(50/120)*(40/100)]=0.025, desconsiderando o denominador P(Tempo Longo)*P(Dispositivo Desktop)*P(Tráfego Orgânico) tendo em vista que irá comparar com a P(Não Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico) com a mesma simplificação.', 2.0)
('Citar e/ou explicitar que, usando Naive Bayes, dado a probabilidade P(Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico)~0.94 ou ~0.094 (simplificando denominador de ambas fórmulas) e P(Não Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico)=0.25 ou 0.025 (simplificando denominador de ambas fórmulas), o sistema apontaria que o cliente irá comprar na plataforma.', 2.0)
==========Resposta do estudante da questão 3:
2) Considerando: Compraram = C, NÃ£o Compraram = NC, Longo = L, Desktop = D e Organico = O
P(C) = 80/200 = 0,4
P(NC) = 120/200 = 0,6
P(L|C) = 60/80 = 0,75
P(L|NC) = 30/120 = 0,25
P(D|C) = 50/80 = 0,625
P(D|NC) = 50/120 = 0,41
P(O|C) = 40/80 = 0,5
P(O|NC) = 40/120 = 0,33
P(C|L&D&O) = P(C)*P(L|C)*P(D|C)*P(O|C)
P(C|L&D&O) = 0,4*0,75*0,625*0,5
P(C|L&D&O) = 0,09375
P(NC|L&D&O) = P(NC)*P(L|NC)*P(D|NC)*P(O|NC)
P(NC|L&D&O) = 0,6*0,25*0,41*0,33
P(NC|L&D&O) = 0,020295
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante apresenta uma abordagem correta na aplicação da fórmula de Naive Bayes para calcular as probabilidades de compra e não compra com base nas características fornecidas.

No entanto, o estudante não normalizou as probabilidades corretamente e não comparou os resultados para tomar uma decisão sobre a compra do cliente. Além disso, o estudante não aplicou a fórmula de Naive Bayes de forma completa, desconsiderando o denominador comum para as probabilidades de compra e não compra.
==========Feedback:
Para melhorar, o estudante deve aplicar a fórmula de Naive Bayes de forma completa, considerando o denominador comum para as probabilidades de compra e não compra. Além disso, o estudante deve comparar os resultados das probabilidades de compra e não compra para tomar uma decisão sobre a compra do cliente. Por exemplo, calculando P(Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico) e P(Não Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico) e comparando os resultados, o estudante pode concluir se o cliente irá comprar ou não. A resposta também deve apresentar os cálculos detalhados e a interpretação dos resultados.
==========Pontuação:
3.0
==========