Nome: GISLAYNE MARIA DA SILVA BATISTA
Pontuação Total da Avaliação: 7.90 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) Em um problema de arvore de decisÃ£o o ganho de informaÃ§Ã£o Ã© uma boa forma de regra de divisÃ£o para classificaÃ§Ã£o, onde ele Ã© essencial para escolher que partiÃ§Ã£o seguir. O ganho de informaÃ§Ã£o tem como conceito fundamento o calculo de Entropia que mede a aleatoriedade das variÃ¡veis. Ou seja, quanto maior a entropia mais misturadas serÃ£o os dados. Por exemplo, Se pegarmos o atributo graduaÃ§Ã£o e calcular o seu ganho de informaÃ§Ã£o atravÃ©s do cÃ¡lculo de entropia obtendo a probabilidade dos candidatos graduados aceitos e rejeitados e os candidatos nÃ£o graduados aceitos e rejeitados e calcular a entropia de cada um (graduado e nÃ£o graduado) e depois calcular a entropia geral do atributo graduaÃ§Ã£o atravÃ©s do seguinte calculo (H(G) = P(CG)*H(CG)+P(SG)*H(SG)) e por fim calcular o ganho de informaÃ§Ã£o atravÃ©s do seguinte calculo (IG(G) = H(I)-H(G)) o resultado obtido vai ser comparado com outros IG dos demais atributos e verificado se esse atributo Ã© ideal ou nÃ£o para a resoluÃ§Ã£o do problema.
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante apresenta uma boa compreensão geral do conceito de entropia e ganho de informação em árvores de decisão. No entanto, falta clareza e detalhes específicos sobre como esses conceitos são aplicados para decidir qual atributo deve ser escolhido como raiz da árvore de decisão. Além disso, não há uma explicação clara sobre o cálculo da entropia inicial de "Aceito" ou "Rejeitado" antes da partição e como o Information Gain (IG) é calculado de forma detalhada para definir o melhor atributo.
==========Feedback:
Para melhorar a resposta, o estudante deve começar explicando o cálculo da entropia inicial de "Aceito" ou "Rejeitado" antes da partição. Em seguida, deve detalhar como o cálculo de Information Gain (IG) é realizado, mostrando que IG = Entropia(Antes da partição) - Entropia(Após partição), e como o maior IG define qual atributo será escolhido. Além disso, é essencial oferecer um exemplo concreto ou fórmula para calcular a entropia final considerando as partições possíveis no atributo analisado. Com esses detalhes, a resposta se tornaria mais completa e rigorosa.
==========Pontuação:
1.6
==========
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) Considerando a entropia inicial como H(I), entropia de experiÃªncia como H(E), Probabilidade dos candidatos Serem aceitos (A), Probabilidade dos candidatos serem Rejeitados como P(R), entropia dos candidatos com experiÃªncia H(CE), entropia candidatos sem experiÃªncia como H(SE) e Ganho de informaÃ§Ã£o como IG(x), sendo x a variÃ¡vel que eu quero calcular.
H(I) = 0,94;
Primeiramente calcular os candidatos com experiÃªncia, total Ã© 49 candidatos (0,49).
P(A)= 0,42/0,49;
P(R)= 0,07/0,49;
H(CE) = -(0,42/0,49)*log2 (0,42/0,49) - (0,07/0,49)*log2(0,07/0,49) =
-(0,857142857)*(â■■0,222392422) - (0,142857143)*(â■■2,807354921) =
0,190622076 + 0,401050703 = 0,591672779 bits.
Agora vamos calcular os candidatos sem experiÃªncia, total Ã© 90 candidatos (0.90).
P(A)= 0.12/0.90;
P(R)= 0.78/0.90;
H(SE) = -(0,12/0,90)*log2 (0,12/0,90) - (0,78/0.90)*log2(0,78/0.90) =
-(0,133333333)*(â■■2,906890599) - (0,866666667)*(â■■0,206450877) =
0,387585412 + 0,178924093 = 0,566509505 bits.
Depois de obter as duas entropias o prÃ³ximo passo Ã© calcular a entropia geral do atributo experiÃªncia.
H(E) = 0,49*0,591672779 + 0,90*0,566509505 =
H(E) = 0,799778216;
H(E) entÃ£o tem o valor aproximadamente 0,80.
Para finalizar iremos calcular o IG (E), ganho de informaÃ§Ã£o do atributo experiencia.
IG(E) = 0,94 - 0,80 =
IG(E) = 0,14;
Podemos analisar pelo resultado obtido atravÃ©s dos cÃ¡lculos que o ganho de informaÃ§Ã£o obtido Ã© mÃ-nimo, caso esse atributo tem um IG menor que outros atributos entÃ£o significa que ele nÃ£o Ã© a escolha ideal para se trabalha com o problema.
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante demonstra um esforço para calcular o ganho de informação do atributo "Experiência" para uma árvore de decisão. No entanto, háprecisão nos cálculos e interpretação dos resultados.
==========Feedback:
O estudante corretamente calculou as entropias para os grupos de candidatos com e sem experiência. A entropia para os candidatos com experiência foi calculada como 0,591672779 bits e para os sem experiência como 0,566509505 bits. O ganho de informação foi calculado como 0,94 - 0,80 = 0,14, o que indica um ganho de informação mínimo. No entanto, a interpretação do

resultado poderia ser mais clara em termos de como o ganho de informação reflete a utilidade do atributo "Experiência" na previsão do resultado. Além disso, os cálculos intermediários poderiam ser apresentados de forma mais clara e organizada.
==========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) Primeiro iremos calcular a probabilidade de se ele compra ou nÃ£o o produto.
P(C) Ã© a probabilidade de comprar e P(N) Ã© a probabilidade de nÃ£o comprar.
P(C) = 80/200 =
P(C) = 0,4;
P(N) = 120/200 =
P(N) = 0,6;
Agora vamos calcular a probabilidade (verosemelhanÃ§a) das caracterÃ-sticas.
P(L/C) Ã© de se o cliente comprou com o tempo na pÃ¡gina sendo curto e P(L/N) Ã© de se o cliente nÃ£o comprou com o tempo na pÃ¡gina sendo curto.
P(L/C) = 0,60/0,40 = 1,5
P(L/N) = 0,30/0,60 = 0,5
P(D/C) Ã© de se o cliente comprou o Dispositivo sendo Desktop e P(D/N) Ã© de se o cliente nÃ£o comprou o Dispositivo sendo Desktop.
P(D/C) = 0,50/0,40 = 1,25
P(D/N) = 0,50/0,60 = 0,833333333
P(O/C) Ã© de se o cliente comprou com a Origem do trÃ¡fego sendo OrgÃ¢nico e P(O/N) Ã© de se o cliente nÃ£o com a Origem do trÃ¡fego sendo OrgÃ¢nico.
P(O/C) = 0,40/0,40 = 1
P(O/N) = 0,40/0,60 = 0,666666667
Agora vamos obter a probabilidade de cada uma das caracterÃ-sticas.
P(L) = 1,5*0,40 + 0,5*0,60 = 0,90

P(D) = 1,25*0,40 + 0,60*0,833333333 = 1
P(O) = 1*0,40 + 0,666666667*0,60 = 0,80
Calculo final P(C/L,D,O) Ã© probabilidade de o cliente comprar sendo o tempo na pÃ¡gina longo, o dispositivo sendo Descktop e a origem do trÃ¡fego sendo OrgÃ¢nico e P(N/L,D,O) Ã© probabilidade do cliente nÃ£o comprar sendo o tempo na pÃ¡gina longo, o dispositivo sendo Descktop e a origem do trÃ¡fego sendo OrgÃ¢nico.
P(C/L,D,O) = (0,40*1,5*1,25*1)/(0,90*1*0,80) =1,041
P(N/L,D,O) = (0,60*0,5*0,83*0,66)/(0,90*1*0,80) = 0,22825
Dado aos resultados obtidos podemos concluir que o cliente tem mais probabilidade de comprar o produto.
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante apresenta uma tentativa de aplicar o algoritmo de Naive Bayes para prever se um cliente comprará ou não um produto com base nas características fornecidas. No entanto, há alguns erros e simplificações que afetam a precisão dos cálculos. O estudante calculou as probabilidades de compra e não compra com base nas características, mas não aplicou corretamente a fórmula de Naive Bayes para obter as probabilidades condicionais.
==========Feedback:
Para atingir a nota máxima, o estudante deveria ter calculado corretamente as probabilidades condicionais usando a fórmula de Naive Bayes. Por exemplo, a probabilidade de compra dado que o tempo na página é longo, o dispositivo é desktop e a origem do tráfego é orgânica pode ser calculada como P(Compra|Tempo Longo, Dispositivo Desktop, Tráfego Orgânico) = [(80/200)*(60/80)*(50/80)*(40/80)] / [(90/200)*(100/200)*(80/180)], o que resulta em uma probabilidade de aproximadamente 0,94. Da mesma forma, a probabilidade de não compra pode ser calculada. Além disso, o estudante deve explicitar claramente os passos do cálculo e justificar a escolha do método de Naive Bayes para esse problema.
==========Pontuação:
4.0
==========