Nome: MATHEUS MUNIZ DE ALBUQUERQUE
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) Para que a empresa consiga fazer uma arvore de decisÃ£o robusta, evitando o overfitting, temos que reduzir a profundidade da arvore, definir um criterio de parada!
Entropia = medimos a inceteza e a desordem dos dados;
Ganho de informaÃ§Ã£o = calcula a reduÃ§Ã£o dessa incerteza da entropia ao dividir os dados com base no atributo;
Tendo em vista isso, o atributo com o maior Ganho de informaÃ§Ã£o, serÃ¡ escolhido como raiz da Ã¡rvore de decisÃ£o, pois ele oferece a melhor separaÃ§Ã£o dos dados, reduzindo a incerteza de forma mais eficiente.
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante aborda a entropia e o ganho de informação ao construir uma árvore de decisão, mas não explicita os cálculos da entropia inicial e final, nem detalha como o ganho de informação é calculado. Além disso, a resposta não apresenta como a entropia é reduzida ao dividir os dados com base nos atributos.
==========Feedback:
Para melhorar, o estudante deve calcular explicitamente a entropia inicial de "Aceito" ou "Rejeitado" antes da partição e explicar como o Information Gain (IG) é calculado como IG = Entropia(Antes da partição) - Entropia(Após partição). Além disso, seria benéfico apresentar o cálculo da entropia final considerando as partições possíveis nos atributos analisados. Isso ajudaria a compreender claramente como o atributo com o maior IG é escolhido como raiz da árvore de decisão.
==========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) entropia inicial = 0,94
Formula entropia:-P(A)xlog2xP(A)-P(B)xlog2xP(B)
candidatos com experiÃªncia:
aceitos:42;
regjeitados:7;
total: 49;
p(aceito) = 42/49 = 0,8571
p(rejeitado) = 7/49 = 0,1428
-0,8571 x log2 x 0,8571 - 0,1428 x log2 x 0,1428 = -0.231
Candidatos sem experiencia:
aceitos:12;
regjeitados:78;
total:90;
p(aceito) = 12/90 = 0,1333
p(rejeitado) = 78/90 = 0,8666
-0,1333xlog2x0,1333-0,8666xlog2x0,8666 = -0.227
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante apresenta cálculos parciais da entropia para os grupos de amostras com e sem experiência, mas não conclui corretamente o cálculo do ganho de informação. O estudante calcula a entropia para os candidatos com experiência como -0,8571 x log2 x 0,8571 - 0,1428 x log2 x 0,1428, obtendo um valor aproximado de 0,59, e para os candidatos sem experiência, calcula -0,1333 x log2 x 0,1333 - 0,8666 x log2 x 0,8666, obtendo um valor aproximado de 0,57. No entanto, o estudante não prossegue com o cálculo do ganho de informação, que é essencial para avaliar a eficácia do atributo "Experiência" na previsão do resultado.
==========Feedback:
Para atingir a nota máxima, o estudante deve completar o cálculo do ganho de informação. O ganho de informação é calculado como a entropia inicial menos a entropia ponderada após a partição. A entropia ponderada é calculada como a soma do produto da probabilidade de cada subconjunto pela sua entropia. Neste caso, a entropia ponderada após a partição "Experiência" é (49/139)*0,59 + (90/139)*0,57. O estudante deve realizar esse cálculo e, em seguida, subtrair o resultado da entropia inicial para obter o ganho de informação. Além disso, o estudante deve interpretar o resultado, explicando o que um ganho de informação maior que zero significa para a homogeneidade dos subconjuntos após a partição.
==========Pontuação:
1.8
==========
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) Para prever se um novo usuario comprarÃ¡ ou nÃ£o um produto apÃ³s visualizar a pÃ¡gina do produto,
levando em consideraÃ§Ã£o as caracteristicas:
Tempo na PÃ¡gina = Longo;
Dispositivo = Desktop;
Origem do TrÃ¡fego = OrgÃ¢nico;
Temos os seguintes calculos:
P(Compra) = 80/200 = 0,4
P(NÃ£o Compra) = 120/200 = 0,6
P(Longo|Compra) = 60/80 = 0,75
P(Longo|NÃ£o Compra) = 30/120 = 0,25
P(Dektop|Compra) = 50/80 = 0,625
P(Dektop|NÃ£o Compra) = 50/120 = 0,5
P(OrgÃ¢nico|Compra) = 40/80 = 0,5
p(OrgÃ¢nico|NÃ£o Compra) = 40/120 = 0,33
P(Longo|Compra) x P(Dektop|Compra) x P(OrgÃ¢nico|Compra) = 0,75 x 0,625 x 0,5 x 0,4 = 0,093
P(Longo|NÃ£o Compra) x P(Dektop|NÃ£o Compra) x P(OrgÃ¢nico|NÃ£o Compra) = 0,25 x 0,5 x 0,33 x 0,6 = 0,024
Logo, a probabilidade de um novo usuario comprar ou nÃ£o comprar Ã©:
Comprar = 0,093/(0,093+0,024) = 0,79(79%)
NÃ£o Comprar = 0,024/(0,024+0,093) = 0,20(20%)
==========
Correção proposta pela IA:
==========
==========Considerações:
A resposta do estudante demonstra um entendimento básico do conceito de Naive Bayes para prever a probabilidade de um cliente comprar ou não um produto com base nas características fornecidas. No entanto, há alguns erros e simplificações que precisam ser corrigidos para atender às rubricas fornecidas.
==========Feedback:
O estudante calculou corretamente as probabilidades iniciais, como P(Compra) e P(Não Compra), e também estimou as probabilidades condicionais, como P(Longo|Compra), P(Dektop|Compra), P(Orgânico|Compra) e suas contrapartes para não compra. No entanto, ao aplicar a fórmula de Naive Bayes, o estudante simplificou incorretamente a equação, multiplicando as probabilidades

condicionais pela probabilidade inicial de compra ou não compra, mas não considerou corretamente o denominador, que é o produto das probabilidades marginais dos atributos. Além disso, a simplificação feita não seguiu exatamente a forma como as rubricas sugerem, o que poderia ter levado a uma solução mais precisa.
==========Pontuação:
4.5
==========