

Nome: GISLAYNE MARIA DA SILVA BATISTA

Pontuação Total da Avaliação: 4.70 pontos

=====

Correção da Questão 1:

Pergunta: 1a) Uma empresa está desenvolvendo um sistema para classificar mensagens recebidas como "Urgente" ou "Não Urgente" com base nas palavras presentes na mensagem. Foi analisado um conjunto de 100 mensagens, e os dados a seguir foram coletados: Mensagens Urgentes: 30; Mensagens Não Urgentes: 70; Palavra Presente "imediatamente": 15 (Mensagens Urgentes) e 5 (Mensagens Não Urgentes); Palavra Presente "problema": 10 (Mensagens Urgentes) e 10 (Mensagens Não Urgentes); Palavra Presente "atraso": 8 (Mensagens Urgentes) e 12 (Mensagens Não Urgentes). Calcule as probabilidades de uma mensagem ser "Urgente" e "Não Urgente" com base no conjunto de dados fornecido. Adicionalmente, determine as probabilidades condicionais para cada palavra ("imediatamente", "problema" e "atraso") em relação às mensagens "Urgentes" e "Não Urgentes".
Diretriz(es): {('Acertar a resposta aproximada de $P(\text{"não urgente"}) = 70/100 = 0.70$, 0.32), ('Acertar a resposta aproximada de $P(\text{"atraso"}|\text{"não urgente"}) = 12/70 \sim 0.17$, 0.32), ('Acertar a resposta aproximada de $P(\text{"urgente"}) = 30/100 = 0.30$, 0.32), ('Acertar a resposta aproximada de $P(\text{"imediatamente"}|\text{"não urgente"}) = 5/70 \sim 0.07$, 0.32), ('Acertar a resposta aproximada de $P(\text{"atraso"}|\text{"urgente"}) = 8/30 \sim 0.27$, 0.32), ('Acertar a resposta aproximada de $P(\text{"problema"}|\text{"urgente"}) = 10/30 \sim 0.33$, 0.32), ('Acertar a resposta aproximada de $P(\text{"problema"}|\text{"não urgente"}) = 10/70 \sim 0.14$, 0.32), ('Acertar a resposta aproximada de $P(\text{"imediatamente"}|\text{"urgente"}) = 15/30 = 0.5$, 0.32)}

=====Resposta do estudante da questão 1:

1A) <<QUESTÃO NÃO RESPONDIDA>>

=====

Correção proposta pela IA:

=====

=====Considerações:

=====Feedback:

=====Pontuação:

0.0

=====

Correção da Questão 2:

Pergunta: 1b) Uma empresa está desenvolvendo um sistema para classificar mensagens recebidas como "Urgente" ou "Não Urgente" com base nas palavras presentes na mensagem. Foi analisado um conjunto de 100 mensagens, e os dados a seguir foram coletados: Mensagens Urgentes: 30; Mensagens Não Urgentes: 70; Palavra Presente "imediatamente": 15 (Mensagens Urgentes) e 5 (Mensagens Não Urgentes); Palavra Presente "problema": 10 (Mensagens Urgentes) e 10 (Mensagens Não Urgentes); Palavra Presente "atraso": 8 (Mensagens Urgentes) e 12 (Mensagens Não Urgentes). Suponha que uma nova mensagem contenha as palavras "imediatamente" e "problema". Calcule a probabilidade de ser uma mensagem "Urgente" e de ser "Não Urgente" utilizando o teorema de Bayes e classifique a mensagem como "Urgente" ou "Não Urgente".

Diretriz(es): {('Classificar o e-mail como "URGENTE" baseado no cálculo do máximo argumento entre $P(\text{"Urgente"}|\text{"imediatamente"}\&\text{"problema"})$ e $P(\text{"Não Urgente"}|\text{"imediatamente"}\&\text{"problema"})$, 0.5), ('Calcular a probabilidade $P(\text{"Urgente"}|\text{"imediatamente"}\&\text{"problema"}) = (P(\text{"urgente"}) * P(\text{"imediatamente"}|\text{"urgente"}) * P(\text{"problema"}|\text{"urgente"})) / (P(\text{"problema"}) * P(\text{"imediatamente"}|\text{"urgente"}))$, podendo simplificar para $P(\text{"Urgente"}|\text{"imediatamente"}\&\text{"problema"}) = P(\text{"urgente"}) * P(\text{"imediatamente"}|\text{"urgente"}) * P(\text{"problema"}|\text{"urgente"}) \sim 0.30 * 0.5 * 0.33 \sim 0.05$, 1.0), ('Calcular a probabilidade $P(\text{"Não Urgente"}|\text{"imediatamente"}\&\text{"problema"}) = (P(\text{"Não urgente"}) * P(\text{"imediatamente"}|\text{"Não urgente"}) * P(\text{"problema"}|\text{"Não urgente"})) / (P(\text{"problema"}) * P(\text{"imediatamente"}|\text{"Não urgente"}))$, podendo simplificar para $P(\text{"Não Urgente"}|\text{"imediatamente"}\&\text{"problema"}) = P(\text{"Não urgente"}) * P(\text{"imediatamente"}|\text{"Não urgente"}) * P(\text{"problema"}|\text{"Não urgente"}) \sim 0.70 * 0.07 * 0.014 \sim 0.001$, 1.0)}

=====Resposta do estudante da questão 2:

1B) <<QUESTÃO NÃO RESPONDIDA>>

=====

Correção proposta pela IA:

=====

=====Considerações:

=====Feedback:

=====Pontuação:

0.0

=====

Correção da Questão 3:

Pergunta: 2a) Árvores de decisão são modelos amplamente utilizados para classificação e regressão. Explique como o conceito de ganho de informação é utilizado na construção de uma árvore de decisão. Detalhe mostrando o uso do cálculo de entropia e ganho de informação em um problema hipotético.

Diretriz(es): {'Explicar que o conceito de Information Gain (IG) como sendo $IG = H(\text{"antes da partição"}) - H(\text{"depois da partição"})$ é essencial para decidir qual atributo/partição escolher a cada nível da árvore', 1.0), ('Demonstrar o cálculo de entropia inicial como sendo $H(\text{"antes da partição"}) = -p(\text{"classe1"}) \log_2(P(\text{"classe1"})) - \dots - p(\text{"classeN"}) \log_2(P(\text{"classeN"}))$ ', 1.0), ('Demonstrar o cálculo de entropia final como sendo $H(\text{"depois da partição"}) = P(\text{"amostras irem para conjunto 1"}) \cdot H(\text{"conjunto 1 após partição"}) + P(\text{"amostras irem para conjunto 2"}) \cdot H(\text{"conjunto 2 após partição"})$ ', 1.0)}

=====Resposta do estudante da questão 3:

2A) O conceito de ganhos de informação é usado para verificar quais atributos são relevantes ao problema por meio do cálculo de entropia. A entropia é usado para calcular a aleatoriedade de variáveis aleatórias e diminuir a entropia da variável alvo para se obter o ganho de informação nesse atributo.

Em uma situação hipotética onde queremos descobrir se vamos jogar naquele dia baseado em uma tabela feitas dos dias onde eles jogaram ou não dado a um conjunto de condições climáticas dos dias.

Levando em consideração que em 14 dias 9 foram dias em que eles jogaram e 5 os que não jogaram é calculado a entropia da coluna joga onde, $H(\text{joga}) = -9/14 \times \log_2(9/14) - 5/14 \times \log_2(5/14)$.

Se escolhermos por exemplo o atributo Clima para calcular a entropia levando em consideração que esse atributo é dividido em três categorias (ensolarado, nublado e chuvoso) sendo a probabilidade de ensolarado ser 5/14, chuvoso ser 5/14 e nublado ser 4/14 iremos calcular a entropia de cada um quando eles jogam e não jogam para obter a entropia final do atributo clima.

$H(\text{joga/ensolarado}) = -3/5 \times \log_2(3/5) - 2/5 \times \log_2(2/5)$

$H(\text{joga/nublado}) = -4/4 \times \log_2(4/4) = 0$

$H(\text{joga/chuvoso}) = -3/5 \times \log_2(3/5) - 2/5 \times \log_2(2/5)$

$H(\text{joga/clima}) = P(\text{ensolarado}) \times H(\text{joga/ensolarado}) + P(\text{nublado}) \times H(\text{joga/nublado}) + P(\text{chuvoso}) \times H(\text{joga/chuvoso})$.

E por último calcular o IG (ganho de informação) que é $H(\text{joga}) - H(\text{joga/clima})$.

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante demonstra uma compreensão básica do conceito de ganho de informação e sua aplicação na construção de árvores de decisão. Ele menciona corretamente que o conceito de ganho de informação é utilizado para verificar quais atributos são relevantes ao problema por meio do cálculo de entropia e que a entropia é usada para calcular a aleatoriedade de variáveis aleatórias. Além disso, ele fornece um exemplo hipotético e calcula a entropia da coluna "joga" e a entropia do atributo "clima" para obter o ganho de informação.

=====Feedback:

Para atingir a nota máxima, o estudante poderia melhorar sua resposta explicando de forma mais detalhada como o conceito de Information Gain (IG) é essencial para decidir qual atributo ou partição escolher a cada nível da árvore. Ele também poderia fornecer mais detalhes sobre o cálculo da entropia inicial e final, e como esses cálculos são utilizados para determinar o ganho de informação. Além disso, seria útil se o estudante explicasse como a escolha do atributo com o maior ganho de informação é feita para criar a árvore de decisão. Em relação ao exemplo hipotético, o estudante poderia fornecer mais detalhes sobre como os cálculos seriam utilizados na prática para determinar se as pessoas jogariam ou não com base nas condições climáticas.

=====Pontuação:

2.2

=====

Correção da Questão 4:

Pergunta: 2b) Árvores de decisão são modelos amplamente utilizados para classificação e regressão. Suponha que você está desenvolvendo um sistema de suporte à decisão, aponte dois critérios de parada que você utilizaria na construção da árvore de decisão para garantir um modelo eficiente?

Diretriz(es): {'Mencionar e/ou explicar dois ou mais critérios de parada válidos na construção de árvores de decisão.', 2.5}, {'Não mencionar e/ou explicar qualquer critério de parada na construção de árvores de decisão.', 0.0}, {'Mencionar e/ou explicar dois apenas um critério de parada válido na construção de árvores de decisão.', 1.25}}

=====Resposta do estudante da questão 4:

2B) Profundidade da Árvore

Homogeneidade de classes

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante menciona dois critérios de parada na construção de árvores de decisão: Profundidade da árvore e Homogeneidade de classes. Esses dois critérios são válidos e importantes para garantir um modelo eficiente.

=====Feedback:

O estudante atingiu parcialmente a descrição da rubrica, pois mencionou dois critérios de parada válidos. No entanto, seria interessante que o estudante explicasse brevemente o porquê da escolha desses critérios e como eles contribuem para a eficiência do modelo. Além disso, uma breve explicação sobre como esses critérios são implementados na prática poderia fortalecer a resposta.

=====Pontuação:

2.5

=====