

Nome: Gislayne Maria

Pontuação Total da Avaliação: 7.11 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): $\{P(\text{"problema"}|\text{"não urgente"})=10/70\sim0.14', 0.32\}$,
 $\{P(\text{"imediatamente"}|\text{"urgente"})=15/30=0.5', 0.32\}$, $\{P(\text{"atraso"}|\text{"não urgente"})=12/70\sim0.17', 0.32\}$, $\{P(\text{"urgente"})=30/100=0.30', 0.32\}$, $\{P(\text{"não urgente"})=70/100=0.70', 0.32\}$,
 $\{P(\text{"problema"}|\text{"urgente"})=10/30\sim0.33', 0.32\}$, $\{P(\text{"atraso"}|\text{"urgente"})=8/30\sim0.27', 0.32\}$,
 $\{P(\text{"imediatamente"}|\text{"não urgente"})=5/70\sim0.07', 0.32\}$

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

1A)

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante não foi fornecida, mas com base nas diretrizes fornecidas, podemos calcular as probabilidades e probabilidades condicionais necessárias. A probabilidade de uma mensagem ser "Urgente" é $30/100 = 0.30$ e a probabilidade de ser "Não Urgente" é $70/100 = 0.70$. Além disso, podemos calcular as probabilidades condicionais para cada palavra em relação às mensagens "Urgentes" e "Não Urgentes".

=====Feedback:

Para melhorar a resposta, o estudante deve fornecer todos os cálculos necessários para as probabilidades e probabilidades condicionais solicitadas. Isso inclui calcular $P(\text{"urgente"})$, $P(\text{"não urgente"})$, $P(\text{"imediatamente"}|\text{"urgente"})$, $P(\text{"imediatamente"}|\text{"não urgente"})$, $P(\text{"problema"}|\text{"urgente"})$, $P(\text{"problema"}|\text{"não urgente"})$, $P(\text{"atraso"}|\text{"urgente"})$ e $P(\text{"atraso"}|\text{"não urgente"})$. Além disso, o estudante deve garantir que todos os cálculos sejam precisos e que as respostas estejam no formato solicitado.

=====Pontuação:

2.3962406015037594

=====

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): $\{P(\text{"Urgente"}|\text{"imediatamente"}\&\text{"problema"})=(P(\text{"urgente"})*P(\text{"imediatamente"}|\text{"urgente"})*P(\text{"problema"}|\text{"urgente"}))/(P(\text{"problema"})*P(\text{"imediatamente"}|\text{"urgente"})+P(\text{"problema"}|\text{"urgente"})*P(\text{"imediatamente"}|\text{"urgente"}))\}$, podendo simplificar para $P(\text{"Urgente"}|\text{"imediatamente"}\&\text{"problema"})=P(\text{"urgente"})*P(\text{"imediatamente"}|\text{"urgente"})*P(\text{"problema"}|\text{"urgente"})\sim0.30*0.5*0.33\sim0.05', 1.0\}$, ('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), ('P("Não Urgente"|"imediatamente"&"problema")=(P("Não urgente")*P("imediatamente"|"Não urgente")*P("problema"|"Não urgente"))/(P("problema")*P("imediatamente"|"Não urgente"))), podendo simplificar para $P(\text{"Não Urgente"}|\text{"imediatamente"}\&\text{"problema"})=P(\text{"Não urgente"})*P(\text{"imediatamente"}|\text{"Não urgente"})$

urgente")*P("problema"|"Não urgente") ~ 0.70 * 0.07 * 0.014 ~ 0.001', 1.0))

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

1B)

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante foi parcialmente correta, pois tentou aplicar o teorema de Bayes para calcular a probabilidade de uma mensagem ser "Urgente" ou "Não Urgente" com base nas palavras presentes. No entanto, a aplicação do teorema foi simplificada e não apresentou os cálculos detalhados para ambas as probabilidades, o que é necessário para uma classificação precisa.

=====Feedback:

Para melhorar a resposta, o estudante deveria calcular as probabilidades condicionais para cada palavra dada a categoria da mensagem ("Urgente" ou "Não Urgente") e então aplicar essas probabilidades no teorema de Bayes. Isso inclui calcular $P(\text{"imediatamente"}|\text{"Urgente"})$, $P(\text{"problema"}|\text{"Urgente"})$, $P(\text{"imediatamente"}|\text{"Não Urgente"})$ e $P(\text{"problema"}|\text{"Não Urgente"})$. Após calcular essas probabilidades, o estudante deveria usar o teorema de Bayes para encontrar $P(\text{"Urgente"}|\text{"imediatamente"} \& \text{"problema"})$ e $P(\text{"Não Urgente"}|\text{"imediatamente"} \& \text{"problema"})$, e então classificar a mensagem com base na probabilidade mais alta. Além disso, deve-se considerar a probabilidade a priori de cada categoria, que é 0,3 para "Urgente" e 0,7 para "Não Urgente".

=====Pontuação:

1.3555555555555556

=====

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

Elevando 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, $H(joga) = -9/14 \times \log_2(9/14) - 5/14 \times \log_2(5/14)$.

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.

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

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

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

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

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

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante demonstra um entendimento básico sobre o conceito de ganho de informação e sua aplicação na construção de árvores de decisão. Ele corretamente identifica a entropia como uma medida de aleatoriedade e reconhece a importância do cálculo de entropia para determinar o ganho de informação.

=====Feedback:

Para melhorar a resposta e alcançar a nota máxima, seria necessário detalhar melhor o conceito de Information Gain (IG) e sua fórmula, $IG = H(\text{'antes da partição'}) - H(\text{'depois da partição'})$, explicando claramente como ele é essencial para decidir qual atributo ou partição escolher a cada nível da árvore. Além disso, exemplificar o cálculo de entropia inicial e final com base em um exemplo concreto, como o problema hipotético apresentado, ajudaria a ilustrar melhor o processo. A resposta também poderia se beneficiar de uma explicação mais clara sobre como as probabilidades são calculadas e utilizadas na fórmula de entropia.

=====Pontuação:

1.9024

=====

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): {'Apontar a limitação de número insuficiente de amostras no nó para realizar nova partição como um possível critério de parada.', 1.25}, {'Apontar a limitação da quantidade de níveis da árvore (profundidade máxima atingida) como um possível critério de parada.', 1.25}, {'Apontar a limitação de entropia das classes (labels) atingir zero no nó como um possível critério de parada.', 1.25}, {'Apontar qualquer outra limitação plausível para que uma árvore de decisão não continue a realizar partições com base em um número de amostras presentes em um dado nó.', 1.25)}

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

2B) Profundidade da Árvore

Homogeneidade de classes

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante aponta dois critérios de parada para a construção de uma árvore de decisão: a profundidade da árvore e a homogeneidade de classes. No entanto, de acordo com as diretrizes fornecidas, esses critérios não correspondem exatamente às opções apresentadas. A profundidade da árvore relaciona-se com a limitação da quantidade de níveis da árvore, o que é um critério de parada, mas a homogeneidade das classes pode ser interpretada como a limitação de entropia das classes atingir zero, embora não seja uma explicação explícita.

=====Feedback:

O estudante demonstrou um entendimento básico ao mencionar a profundidade da árvore, que é um critério de parada relacionado à limitação da quantidade de níveis da árvore. No entanto, para atingir a nota máxima, seria necessário mencionar explicitamente e com detalhes ao menos dois critérios de parada, como a limitação do número insuficiente de amostras no nó para realizar nova partição, a limitação da quantidade de níveis da árvore (profundidade máxima atingida), ou a limitação de entropia das classes atingir zero no nó. A resposta poderia ser melhorada com exemplos concretos ou explicações mais detalhadas sobre como esses critérios são aplicados na prática.

=====Pontuação:

1.4594594594594594

=====