

Nome: Alysson Henrique

Pontuação Total da Avaliação: 6.34 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): $\{('P("urgente")=30/100=0.30', 0.32), ('P("atraso"|"não urgente")=12/70\sim0.17', 0.32), ('P("imediatamente"|"urgente")=15/30=0.5', 0.32), ('P("imediatamente"|"não urgente")=5/70\sim0.07', 0.32), ('P("atraso"|"urgente")=8/30\sim0.27', 0.32), ('P("problema"|"urgente")=10/30\sim0.33', 0.32), ('P("não urgente")=70/100=0.70', 0.32), ('P("problema"|"não urgente")=10/70\sim0.14', 0.32)\}$

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

1A) Probabilidade de mensagem urgente ($P(U)$) = $30/30+70 = 0,3$
Probabilidade de mensagem não urgente ($P(NU)$) = $70/30+70 = 0,7$
Probabilidade da palavra imediatamente $P(IM) = 20/(30+70)$
Probabilidade da palavra problema $P(P) = 20/(30+70)$
Probabilidade da palavra atraso $P(A) = 20/(30+70)$
Probabilidade de mensagem urgente ter palavra imediatamente $[P(U|IM)] = 0,5$
Probabilidade de mensagem não urgente ter palavra imediatamente $[P(NU|IM)] = 0,07$
Probabilidade de mensagem urgente ter palavra problema $[P(U|P)] = 0,33$
Probabilidade de mensagem não urgente ter palavra problema $[P(NU|P)] = 0,14$
Probabilidade de mensagem urgente ter palavra atraso $[P(U|A)] = 0,26$
Probabilidade de mensagem não urgente ter palavra atraso $[P(NU|A)] = 0,17$
Probabilidade de imediatamente estar em urgente = $[P(IM|U)]$
 $P(IM|U) = P(IM)*P(U|IM) / P(U) = 0,33$
Probabilidade de imediatamente não estar em urgente = $[P(IM|NU)]$
 $P(IM|NU) = P(IM)*P(NU|IM) / P(NU) = 0,02$
Probabilidade de problema estar em urgente = $[P(P|U)]$
 $P(P|U) = P(P) * P(U|P) / P(U) = 0,22$
Probabilidade de problema não estar em urgente = $[P(P|NU)]$
 $P(P|NU) = P(P) * P(NU|P) / P(NU) = 0,04$
Probabilidade de atraso estar em urgente = $[P(A|U)]$
 $P(A|U) = P(A) * P(U|A) / P(U) = 0,17$
Probabilidade de atraso não estar em urgente = $[P(A|NU)]$
 $P(A|NU) = P(A) * P(NU|A) / P(NU) = 0,04$

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante apresenta alguns erros e falta de clareza em relação às diretrizes fornecidas. As probabilidades iniciais de mensagens urgentes e não urgentes estão corretas, mas as probabilidades das palavras "imediatamente", "problema" e "atraso" não são calculadas corretamente. Além disso, as probabilidades condicionais também apresentam erros de cálculo.

=====Feedback:

Um feedback detalhado para o estudante seria: primeiro, calcule as probabilidades das palavras "imediatamente", "problema" e "atraso" corretamente, considerando a frequência de cada palavra em relação ao total de mensagens. Em seguida, recalcule as probabilidades condicionais usando a fórmula correta. É importante também explicitar os passos de cálculo e justificar as respostas com base nas diretrizes fornecidas.

=====Pontuação:

1.7455045871559631

=====

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

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

1B) Probabilidade de mensagem que tem â■■■imediatamenteâ■■■ e â■■■problemaâ■■■ ser urgente = $P(IM;P|U)$

$P(IM;P|U) = P(IM|U)*P(P|U)*P(U) = 0,02$

Probabilidade de mensagem que tem â■■■imediatamenteâ■■■ e â■■■problemaâ■■■ não ser urgente = $P(IM;P|NU)$

$P(IM;P|NU) = P(IM|NU)*P(P|NU)*P(NU) = 0,00005$

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante demonstra uma tentativa de aplicar o teorema de Bayes para calcular as probabilidades, mas apresenta erros significativos na formulação e nos cálculos. As diretrizes fornecem uma abordagem correta para resolver o problema, destacando a necessidade de calcular as probabilidades condicionais corretas e aplicar o teorema de Bayes de forma adequada.

=====Feedback:

Um ponto forte na resposta do estudante é o reconhecimento da necessidade de aplicar o teorema de Bayes. No entanto, os cálculos realizados estão incorretos. Para melhorar, o estudante deve entender que as probabilidades condicionais, como $P("imediatamente"|"urgente")$ e $P("problema"|"urgente")$, precisam ser calculadas com base nos dados fornecidos. Além disso, a aplicação do teorema de Bayes requer uma compreensão clara das fórmulas e dos conceitos por trás delas. A resposta do estudante não contempla a normalização necessária para calcular as probabilidades finais de forma correta. Mais precisamente, para calcular $P("Urgente"|"imediatamente"&"problema")$, o estudante deve seguir as diretrizes fornecidas, simplificando a fórmula do teorema de Bayes para o problema específico e calculando cada termo com base nos dados fornecidos. Da mesma forma, para $P("Não Urgente"|"imediatamente"&"problema")$, os cálculos devem refletir as probabilidades condicionais e a priori de acordo com as diretrizes.

=====Pontuação:

0.6311475409836066

=====

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): $\{('Demonstrar o cálculo de entropia final como sendo } H("depois da$

partição")=P("amostras irem para conjunto 1")*H("conjunto 1 após partição")+P("amostras irem para conjunto 2")*H("conjunto 2 após partição"), 1.0), ('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))

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

2A) O ganho de informação é utilizado para que determinar qual atributo ou partição o mais adequado para ser um ponto de decisão da árvore, suponha hipoteticamente que existe um banco de dados sobre jogos de vôlei nesse conjunto temos uma tabela que marca se as pessoas jogaram ou não naquele dia, desses valores temos 8/14 joga e 6/14 não joga, outros dois atributos dessa tabela são, tempo = [nublado: 7/14 joga, limpo 7/14 joga] e feriado = [não: 0/14 joga, sim: 14/14 joga], ao realizar os cálculos temos que a entropia inicial dada pelo cálculo $[-(8/14)\log_2(8/14)-(6/14)\log_2(6/14)]$ será aproximadamente 0,97, aplicando o método para calcular as entropias do atributo tempo = 1 e feriado = 0, podemos fazer o cálculo do ganho de conhecimento para cada atributo subtraindo sua entropia da inicial e verificando que feriado é o atributo melhor para iniciar a árvore de decisão.

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante demonstra uma compreensão básica do conceito de ganho de informação e entropia na construção de árvores de decisão. Ele apresentou um exemplo hipotético sobre um banco de dados de jogos de vôlei e calculou a entropia inicial, mostrando como ela é utilizada para determinar a melhor partição. No entanto, a explicação poderia ser mais detalhada e clara em relação às fórmulas e conceitos utilizados.

=====Feedback:

Um ponto forte da resposta é a apresentação de um exemplo concreto, o que facilita a compreensão do conceito. No entanto, para melhorar, o estudante deveria explicitar as fórmulas de entropia e ganho de informação de forma mais clara e detalhada. Além disso, a explicação do cálculo da entropia final após a partição e a utilização do conceito de Information Gain (IG) poderia ser mais precisa e seguir a fórmula $IG=H(\text{"antes da partição"})-H(\text{"depois da partição"})$.

=====Pontuação:

1.5785046728971963

=====

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 da quantidade de níveis da árvore (profundidade máxima atingida) 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), ('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 de entropia das classes (labels) atingir zero no nó como um possível critério de parada.', 1.25)}

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

2B) Primeiro critério de parada seria determinar uma profundidade máxima para a árvore, outro critério de parada seria determinar um valor como limite de pureza ou entropia que uma folha pode obter e finalizar o processo de criação da árvore quando esse valor for atingido, esses dois métodos permitem gerar árvores de decisão mais eficientes pois impedem que elas se tornem desnecessariamente longas.

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante apresenta uma boa compreensão dos critérios de parada para a construção

de árvores de decisão, mas pode ser aprimorada. O estudante menciona dois critérios de parada importantes: a profundidade máxima da árvore e o valor de entropia ou pureza das classes (labels) em um nó. No entanto, a resposta pode ser mais completa e alinhada às diretrizes fornecidas.

=====Feedback:

Um ponto forte da resposta é a menção à profundidade máxima como um critério de parada, o que está alinhado com uma das diretrizes. Além disso, o estudante menciona o valor de entropia ou pureza como outro critério, o que é relação com a diretriz sobre a limitação da entropia das classes atingir zero. No entanto, para melhorar e atingir a nota máxima, o estudante poderia ter mencionado explicitamente a limitação de número insuficiente de amostras no nó como um possível critério de parada, o que é outra diretriz importante. Além disso, explicitar melhor como esses critérios contribuem para a eficiência do modelo também ajudaria.

=====Pontuação:

2.3807692307692307

=====