

Nome: ALYSSON HENRIQUE DE SOUZA PINHEIRO

Pontuação Total da Avaliação: 8.10 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 $P(\text{"problema"}|\text{"urgente"})=10/30\sim0.33$ ', 0.32}, ('Acertar a resposta aproximada de $P(\text{"não urgente"})=70/100=0.70$ ', 0.32), ('Acertar a resposta aproximada de $P(\text{"atraso"}|\text{"urgente"})=8/30\sim0.27$ ', 0.32), ('Acertar a resposta aproximada de $P(\text{"problema"}|\text{"não urgente"})=10/70\sim0.14$ ', 0.32), ('Acertar a resposta aproximada de $P(\text{"urgente"})=30/100=0.30$ ', 0.32), ('Acertar a resposta aproximada de $P(\text{"imediatamente"}|\text{"urgente"})=15/30=0.5$ ', 0.32), ('Acertar a resposta aproximada de $P(\text{"atraso"}|\text{"não urgente"})=12/70\sim0.17$ ', 0.32), ('Acertar a resposta aproximada de $P(\text{"imediatamente"}|\text{"não urgente"})=5/70\sim0.07$ ', 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 contém alguns erros e imprecisões em relação às probabilidades calculadas. É importante revisar os cálculos e considerar as definições corretas de probabilidade condicional para cada palavra em relação às mensagens "Urgentes" e "Não Urgentes".

=====Feedback:

O estudante calculou corretamente as probabilidades básicas de uma mensagem ser "Urgente" ($P(U) = 0,3$) e "Não Urgente" ($P(NU) = 0,7$), o que é um ponto forte. No entanto, os cálculos das probabilidades condicionais para cada palavra ("imediatamente", "problema" e "atraso") em relação às mensagens "Urgentes" e "Não Urgentes" precisam ser revistos. Por exemplo, a

probabilidade de uma mensagem "Urgente" conter a palavra "problema" é de fato $10/30 = 0,33$, e a probabilidade de uma mensagem "Não Urgente" conter a palavra "problema" é $10/70 = 0,14$, o que está de acordo com as rubricas fornecidas. Já a probabilidade de uma mensagem "Urgente" conter a palavra "atraso" é $8/30 = 0,27$, e a probabilidade de uma mensagem "Não Urgente" conter a palavra "atraso" é $12/70 = 0,17$, conforme solicitado. As probabilidades condicionais, como $P(\text{"imediatamente"}|\text{"urgente"})$, $P(\text{"problema"}|\text{"urgente"})$, e $P(\text{"atraso"}|\text{"urgente"})$, precisam ser calculadas usando a fórmula de probabilidade condicional correta, que é $P(A|B) = P(A \text{ e } B) / P(B)$.

=====Pontuação:

2.2

=====

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): {'Calcular a probabilidade $P(\text{"Não Urgente"}|\text{"imediatamente"}\&\text{"problema"}) = (P(\text{"Não urgente"}) * P(\text{"imediatamente"}|\text{"Não urgente"}) * P(\text{"problema"}|\text{"Não urgente"})) / (P(\text{"problema"}) * P(\text{"imediatamente"}))$, 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), ('Calcular a probabilidade $P(\text{"Urgente"}|\text{"imediatamente"}\&\text{"problema"}) = (P(\text{"urgente"}) * P(\text{"imediatamente"}|\text{"urgente"}) * P(\text{"problema"}|\text{"urgente"})) / (P(\text{"problema"}) * P(\text{"imediatamente"}))$, 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), ('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)}

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

1B) Probabilidade de mensagem que tem "imediatamente" e "problema" 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 "imediatamente" e "problema" 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 foi avaliada considerando as rubricas fornecidas. O estudante tentou aplicar o teorema de Bayes para calcular as probabilidades, mas houve erros significativos nos cálculos e interpretações. Primeiramente, é importante notar que o estudante não seguiu corretamente a fórmula do teorema de Bayes para calcular as probabilidades condicionais. Além disso, os cálculos apresentados foram realizados de maneira simplificada e sem considerar a probabilidade a priori de cada evento e a probabilidade das palavras dadas esses eventos.

=====Feedback:

Para alcançar a nota máxima, o estudante deveria ter calculado as probabilidades de maneira mais precisa e detalhada. Primeiro, calcular a probabilidade de uma mensagem ser urgente dado que contém as palavras "imediatamente" e "problema" usando a fórmula $P(U|IM;P) = P(IM|U) * P(P|U) * P(U) / P(IM;P)$. Da mesma forma, calcular a probabilidade de uma mensagem ser não urgente dado que contém essas palavras usando $P(NU|IM;P) = P(IM|NU) * P(P|NU) * P(NU) / P(IM;P)$. Além disso, a classificação da mensagem como "Urgente" ou "Não Urgente" depende de qual probabilidade condicional é maior.

=====Pontuação:

1.2

=====

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 $H(\text{"depois da partição"}) = P(\text{"amostras irem para conjunto 1"}) * H(\text{"conjunto 1 após partição"}) + P(\text{"amostras irem para conjunto 2"}) * H(\text{"conjunto 2 após partição"})$ ', 1.0)}

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

2A) O ganho de informação é utilizado para que determinar qual atributo ou partição é 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 aborda o conceito de ganho de informação e sua aplicação na construção de árvores de decisão. Ele fornece um exemplo hipotético de um problema de classificação, onde as pessoas jogam ou não jogam vôlei, e como o cálculo de entropia e ganho de informação é utilizado para determinar o melhor atributo para iniciar a árvore de decisão. No entanto, a resposta pode ser melhorada com mais detalhes sobre a fórmula de cálculo de entropia e ganho de informação.

=====Feedback:

Um ponto forte da resposta é a apresentação de um exemplo concreto para ilustrar o conceito de ganho de informação. No entanto, para atingir a nota máxima, o estudante precisaria fornecer mais detalhes sobre a fórmula de cálculo de entropia e ganho de informação, explicitando como ela é usada para decidir qual atributo ou partição é o mais adequado a cada nível da árvore. Além disso, seria útil uma explicação mais clara sobre como o ganho de informação é calculado para cada atributo e como isso determina a escolha do atributo para a próxima partição.

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

=====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 menciona dois critérios de parada válidos para a construção de árvores de decisão, quais sejam, a profundidade máxima da árvore e o limite de pureza ou entropia que uma folha pode obter. Esses critérios são explicados de forma clara e concisa, demonstrando um entendimento básico do assunto.

=====Feedback:

O estudante apresentou uma resposta clara e objetiva, o que é um ponto forte. No entanto, para melhorar, o estudante poderia fornecer mais detalhes sobre como esses critérios são implementados na prática e como afetam a eficiência do modelo. Além disso, exemplos práticos ou ilustrações poderiam ajudar a tornar a resposta mais didática e interessante.

=====Pontuação:

2.5

=====