

Nome: AlexsandroCostaDaSilva

Pontuação Total da Avaliação: 5.77 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) U (Urgente)  $\hat{=}$  30/100 = 0,3

NU (NÃO urgente)  $\hat{=}$  70/100 = 0,7

Probabilidades Condicionais:

$\hat{=}$  imediatamente  $\hat{=}$  20/100 = 0,2

$\hat{=}$  problema  $\hat{=}$  20/100 = 0,2

$\hat{=}$  atraso  $\hat{=}$  20/100 = 0,2

$P(\hat{=}$  imediatamente  $\hat{=}$  U) = 15/20 = 0,75

$P(\hat{=}$  imediatamente  $\hat{=}$  NU) = 5/20 = 0,25

$P(\hat{=}$  problema  $\hat{=}$  U) = 10/20 = 0,5

$P(\hat{=}$  problema  $\hat{=}$  NU) = 10/20 = 0,5

$P(\hat{=}$  atraso  $\hat{=}$  U) = 8/20 = 0,4

$P(\hat{=}$  atraso  $\hat{=}$  NU) = 15/20 = 0,6

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante apresenta alguns erros e falta de clareza em relação às probabilidades condicionais. As probabilidades de uma mensagem ser "Urgente" e "Não Urgente" estão corretas, calculadas como 30/100 = 0,3 e 70/100 = 0,7, respectivamente. No entanto, as probabilidades condicionais para cada palavra apresentam erros de cálculo.

=====Feedback:

Um ponto forte da resposta é o cálculo inicial das probabilidades de mensagens serem "Urgentes" ou "Não Urgentes". Entretanto, os cálculos das probabilidades condicionais para as palavras "imediatamente", "problema" e "atraso" apresentam erros. Por exemplo, para "imediatamente", a probabilidade condicional dado que a mensagem é urgente deve ser calculada como  $P(\text{"imediatamente"}|U) = 15/30$ , e não como foi apresentado. De forma semelhante,  $P(\text{"problema"}|U) = 10/30$ ,  $P(\text{"atraso"}|U) = 8/30$ , e assim por diante. É importante notar que as probabilidades condicionais devem considerar a totalidade das mensagens "Urgentes" ou "Não Urgentes" para cada cálculo específico. Também há uma confusão no cálculo geral das probabilidades das palavras, que devem ser calculadas com base nos totais específicos de cada categoria de mensagem.

=====Pontuação:

1.94112

=====

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)  $P(U|\hat{\text{imediatamente}}) = (0,3*0,75)/0,2 = 1,125$

$P(NU|\hat{\text{imediatamente}}) = (0,3*0,25)/0,2 = 0,375$

$P(U|\hat{\text{problema}}) = (0,3*0,5)/0,2 = 0,75$

$P(NU|\hat{\text{problema}}) = (0,7*0,5)/0,2 = 1,75$

$P(U|\hat{\text{imediatamente}} \text{ e } \hat{\text{problema}}) = 0,3*1,125*0,75 = 0,25$

$P(NU|\hat{\text{imediatamente}} \text{ e } \hat{\text{problema}}) = 0,7*0,375*1,75 = 0,46$

A mensagem é classificada como  $\hat{\text{Não Urgente}}$ , pois a maior probabilidade foi a de  $P(NU|\hat{\text{imediatamente}} \text{ e } \hat{\text{problema}}) = 0,46$ .

=====Correção proposta pela IA:

=====Considerações:

A resposta do estudante apresenta alguns erros de cálculo e interpretação. Em primeiro lugar, o estudante não aplicou corretamente o teorema de Bayes para calcular as probabilidades condicionais. Além disso, os cálculos realizados não levaram em consideração as probabilidades a priori e a posteriori corretas.

=====Feedback:

O estudante precisou calcular as probabilidades condicionais de uma mensagem ser urgente ou não urgente dado que contém as palavras "imediatamente" e "problema". Para isso, deveria ter aplicado o teorema de Bayes considerando as probabilidades a priori de uma mensagem ser urgente ( $P(U)$ ) e não urgente ( $P(NU)$ ), bem como as probabilidades condicionais de uma mensagem conter "imediatamente" e "problema" dado que é urgente ( $P("imediatamente"|U)$  e  $P("problema"|U)$ ) e não urgente ( $P("imediatamente"|NU)$  e  $P("problema"|NU)$ ). Os cálculos apresentados não refletem essas probabilidades de forma correta.

=====Pontuação:

0.875257731958763

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

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

2A) Ganho de informação é o resultado da diferença da entropia inicial calculada para a entropia de alguma variável para melhorar o resultado da classificação. O exemplo hipotético seria para dizer se uma turma de pessoa pode jogar no dia, dando as variáveis como

clima, vento, umidade e outras. Começando com o cálculo da entropia inicial, que seria contar todos os casos, separar os que de jogar são verdadeiros e os que são falsos com a seguinte fórmula  $H(\text{Joga}) = -P(\text{verdadeiro})\log_2(P(\text{verdadeiro})) - P(\text{falso})\log_2(P(\text{falso}))$ . Depois disso poderia ser analisado os casos em que estava ventando separar quando os casos verdadeiros e falsos e calcular a entropia de cada um para quando o Jogar for verdadeiro e falso com as seguintes fórmulas  $H(\text{Joga}|\text{VentoVerdadeiro}) = -P(\text{JogaVerdadeiro})\log_2(P(\text{JogaVerdadeiro})) - P(\text{JogaFalso})\log_2(P(\text{JogaFalso}))$  e  $H(\text{Joga}|\text{VentoFalso}) = -P(\text{JogaVerdadeiro})\log_2(P(\text{JogaVerdadeiro})) - P(\text{JogaFalso})\log_2(P(\text{JogaFalso}))$ , depois de ter os dois resultados aplica-los na seguinte fórmula  $H(\text{Vento}) = P(\text{VentoVerdadeiro}) * H(\text{Joga}|\text{VentoVerdadeiro}) + P(\text{VentoFalso}) * H(\text{Joga}|\text{VentoFalso})$  e com o resultado disso podemos calcular o ganho de informação usando  $H(\text{Joga}) - H(\text{Vento})$ .

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante aborda o conceito de ganho de informação e entropia na construção de uma árvore de decisão, mas carece de clareza e precisão em alguns pontos. O estudante tenta explicar como o cálculo de entropia inicial é feito e como o ganho de informação é utilizado para decidir a partição da árvore. No entanto, a explicação poderia ser mais clara e detalhada, especialmente em relação às fórmulas utilizadas.

=====Feedback:

Uma área que necessita de melhoria é a explicação do conceito de Information Gain (IG) e como ele é calculado. O estudante menciona que o ganho de informação é o resultado da diferença da entropia inicial e a entropia de alguma variável, mas não explica claramente como isso é feito. Além disso, a fórmula para o cálculo da entropia final após a partição poderia ser explicitada de forma mais clara. Para atingir a nota máxima, o estudante deveria demonstrar uma compreensão mais sólida dos conceitos de entropia e ganho de informação e explicá-los de forma clara e precisa.

=====Pontuação:

1.7078260869565218

=====

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) Os critérios de paradas usados seriam o de profundidade máxima e o de quantidade limite por folhas, usando um desses dois modelos é garantido que o sistema não passe do limite esperado e pare quando estiver em uma das condições sem perder a pureza.

=====

Correção proposta pela IA:

=====

=====Considerações:

A resposta do estudante apresenta uma boa intenção, mas carece de precisão e profundidade em relação às diretrizes fornecidas. O estudante menciona dois critérios de parada, sendo eles a profundidade máxima e a quantidade limite por folhas, o que está parcialmente alinhado com as diretrizes.

=====Feedback:

Para atingir a nota máxima, seria necessário que o estudante apresentasse critérios de parada mais específicos e relacionados às diretrizes, como a limitação de número insuficiente de

amostras no nó para realizar nova partição, a limitação da quantidade de níveis da árvore (profundidade máxima atingida) e a limitação de entropia das classes (labels) atingir zero no nó. Além disso, uma explicação mais detalhada sobre como esses critérios contribuem para um modelo eficiente seria benéfica.

=====Pontuação:

1.25

=====