Parte superior do formulário

**Questão 1**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Considere que o departamento de RH,  por meio da sua área de People Analytics, tenha o interesse em investigar as variáveis explicativas  que influenciam na probabilidade de um trainee chegar atrasado à empresa. Sendo assim, o RH elaborou uma pesquisa com os 100 trainees perguntando se ele costuma ou não chegar atrasado no trabalho. Perguntou também sobre a distância percorrida no trajeto (em quilômetros), o número de semáforos pelos quais cada um passou, o período em que foi realizado o trajeto (manhã ou tarde) e como cada um se considera em termos de perfil ao volante (calmo, moderado ou agressivo).

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| tempo | Tempo para chegar à empresa (em minutos) |  |
|  |
| dist | Distância percorrida até a empresa (quilômetros) |  |
|  |
| sem | Quantidade de semáforos |  |
|  |
| per | Período do dia no qual o trainee foi à empresa. Manhã (dummy = 1): caso o trainee tenha ido mais cedo para atividades específicas de sua área ou Tarde (dummy = 0): se o trainee foi apenas para o período da tarde, horário definido para as atividades pertinentes ao programa de Trainee |  |
|  |
|  |
| perfil | Perfil ao volante (Perfil 1: calmo, Perfil 2: moderado e Perfil 3: agressivo) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de atraso do trainee - covariáveis: dist, sem, per e perfil.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | -30,2003 | 9,9811 | -3,03 | 0,002 |  |
| dist | 0,2202 | 0,1097 | 2,01 | 0,045 |  |
| sem | 2,7667 | 0,9217 | 3,00 | 0,003 |  |
| per | -3,6534 | 0,8781 | -4,16 | <0,001 |  |
| perfil\_2 | 1,3460 | 0,7477 | 1,80 | 0,072 |  |
| perfil\_3 | 2,91447 | 1,1788 | 2,47 | 0,013 |  |

Número de observações = 100

Output obtido via stepwise do modelo de Regressão Logística estimado para o problema de atraso do trainee - covariáveis: dist, sem, per e perfil.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | -30,9334 | 10,6363 | -2,91 | 0,004 |  |
| dist | 0,2041 | 0,1012 | 2,02 | 0,044 |  |
| sem | 2,9201 | 1,0108 | 2,89 | 0,004 |  |
| per | -3,7763 | 0,8467 | -4,46 | <0,001 |  |
| perfil\_3 | 2,45907 | 1,1395 | 2,16 | 0,031 |  |

Qual das equações abaixo representa corretamente o cálculo da probabilidade média de chegar atrasado?

Escolha uma:

a. 

b. 

c. 

d. 

**Feedback**

Sua resposta está correta.

A resposta correta é: .

**Questão 2**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Uma empresa de concessão de crédito tem o intuito de avaliar a probabilidade de que seus clientes não cumpram com seus compromissos de pagamentos (probabilidade de default). Por meio de uma base de dados com 2.000 clientes que obtiveram crédito recentemente, a empresa pretende ajustar um modelo de regressão logística binária, utilizando como variáveis explicativas a idade, o sexo (feminino = 0; masculino = 1) e a renda mensal (R$) de cada indivíduo. A variável dependente se refere ao default propriamente dito (não default = 0; default = 1).

Output do modelo de Regressão Logística estimado para o problema de concessão de crédito - covariáveis: idade, sexo e renda.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| Intercepto | 2,9751 | 0,2623 | 11,34 | < 0,001 |  |
| idade | -0,0243 | 0,0070 | -3,49 | < 0,001 |  |
| sexo\_masc | 0,7415 | 0,1135 | 6,53 | < 0,001 |  |
| renda | -0,00026 | 0,0000 | -15,03 | < 0,001 |  |

Número de observações = 2000

Com relação à interpretação da variável sexo, escolha a alternativa correta:

Escolha uma:

a. Como o valor do parâmetro estimado para a variável sexo é maior do que 0, os indivíduos do sexo feminino (dummy = 0) apresentam maiores probabilidades de default do que os do sexo masculino mantidas as demais condições constantes.

b. Como o valor do parâmetro estimado para a variável sexo é menor do que 1, os indivíduos do sexo feminino (dummy = 0) apresentam maiores probabilidades de default do que os do sexo masculino, mantidas as demais condições constantes.

c. Como o sinal do parâmetro estimado para a variável sexo é positivo, os indivíduos do sexo feminino (dummy = 0) apresentam menores probabilidades de default do que os do sexo masculino, mantidas as demais condições constantes.

d. O sexo masculino apresenta uma probabilidade de default equivalente a 0,77415.

**Feedback**

Sua resposta está correta.

A resposta correta é: Como o sinal do parâmetro estimado para a variável sexo é positivo, os indivíduos do sexo feminino (dummy = 0) apresentam menores probabilidades de default do que os do sexo masculino, mantidas as demais condições constantes..

**Questão 3**

Incorreto

Atingiu 0,00 de 1,00

Marcar questão

**Texto da questão**

Considere que o departamento de RH,  por meio da sua área de People Analytics, tenha o interesse em investigar as variáveis explicativas  que influenciam na probabilidade de um trainee chegar atrasado à empresa. Sendo assim, o RH elaborou uma pesquisa com os 100 trainees perguntando se ele costuma ou não chegar atrasado no trabalho. Perguntou também sobre a distância percorrida no trajeto (em quilômetros), o número de semáforos pelos quais cada um passou, o período em que foi realizado o trajeto (manhã ou tarde) e como cada um se considera em termos de perfil ao volante (calmo, moderado ou agressivo).

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| tempo | Tempo para chegar à empresa (em minutos) |  |
|  |
| dist | Distância percorrida até a empresa (quilômetros) |  |
|  |
| sem | Quantidade de semáforos |  |
|  |
| per | Período do dia no qual o trainee foi à empresa. Manhã (categoria de referência - dummy = 1): caso o trainee tenha ido mais cedo para atividades específicas de sua área ou Tarde (dummy = 0): se o trainee foi apenas para o período da tarde, horário definido para as atividades pertinentes ao programa de Trainee |  |
|  |
|  |
| perfil | Perfil ao volante (Perfil 1: calmo, Perfil 2: moderado e Perfil 3: agressivo) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de atraso do trainee - covariáveis: dist, sem, per e perfil.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | -30,2003 | 9,9811 | -3,03 | 0,002 |  |
| dist | 0,2202 | 0,1097 | 2,01 | 0,045 |  |
| sem | 2,7667 | 0,9217 | 3,00 | 0,003 |  |
| per | -3,6534 | 0,8781 | -4,16 | <0,001 |  |
| perfil\_2 | 1,3460 | 0,7477 | 1,80 | 0,072 |  |
| perfil\_3 | 2,91447 | 1,1788 | 2,47 | 0,013 |  |

Número de observações = 100

Output obtido via stepwise do modelo de Regressão Logística estimado para o problema de atraso do trainee - covariáveis: dist, sem, per e perfil.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | -30,9334 | 10,6363 | -2,91 | 0,004 |  |
| dist | 0,2041 | 0,1012 | 2,02 | 0,044 |  |
| sem | 2,9201 | 1,0108 | 2,89 | 0,004 |  |
| per | -3,7763 | 0,8467 | -4,46 | <0,001 |  |
| perfil\_3 | 2,45907 | 1,1395 | 2,16 | 0,031 |  |

Considerando as covariáveis presentes no modelo, qual o perfil menos propenso a chegar atrasado na empresa?

Escolha uma:

a. Trainees que moram mais longe, passam por uma quantidade maior de semáforos, fazem o trajeto na parte da manhã e que apresentam perfil na direção calmo ou moderado.

b. Trainees que moram mais longe, passam por uma quantidade menor de semáforos, fazem o trajeto na parte da manhã e que apresentam perfil na direção calmo ou moderado.

c. Trainees que moram mais perto, passam por uma quantidade menor de semáforos, fazem o trajeto na parte da manhã e que apresentam perfil na direção calmo ou moderado.

d. Trainees que moram mais perto, passam por uma quantidade menor de semáforos, fazem o trajeto na parte da tarde e que apresentam perfil na direção calmo ou moderado.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: Trainees que moram mais perto, passam por uma quantidade menor de semáforos, fazem o trajeto na parte da manhã e que apresentam perfil na direção calmo ou moderado..

**Questão 4**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Considere a área de Digital Analytics de uma empresa varejista que está iniciando sua operação no e-commerce, e deseja  melhorar  a eficácia de sua publicidade a partir de impressões de seus anúncios baseado no perfil do usuário. Para isso, está avaliando diversas informações coletadas a partir da navegação web, e deseja estimar a probabilidade de clique do usuário no anúncio apresentado.

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| Idade | Idade do usuário em anos |  |
|  |
| Renda\_região | Renda média da região na qual o usuário está localizado |  |
|  |
| Gênero | 1, se masculino e 0, se feminino |  |
|  |
| Tempo\_internet | Tempo gasto por dia na internet (em minutos) - obtido a partir dos dados da DMP (Data Management Platform) |  |
|  |
| Tempo\_site | Tempo gasto por dia no site da varejista (em minutos) - obtido com o antigo pixel instalado no site |  |
|  |
| Clicou | Flag se o usuário clicou no anúncio (1) ou não clicou (0) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de clique no anúncio - covariáveis: Idade, Renda\_região, Gênero, Tempo\_internet e Tempo\_site.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| (Intercept) | 28,51771 | 3,49556 | 8158,00000 | <0,001 |  |
| Tempo\_site | -0,20172 | 0,02647 | -7,62100 | <0,001 |  |
| Idade | 0,17601 | 0,03093 | 5,69000 | <0,001 |  |
| Renda\_região | -0,00013 | 0,00002 | -6,24500 | <0,001 |  |
| Tempo\_internet | -0,06931 | 0,00896 | -7,73900 | <0,001 |  |
| Gênero | -0,20946 | 0,50116 | -0,41800 | 0,6760 |  |

Número de observações = 700.

Output obtido após o processo de redução de variáveis do modelo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| Intercepto | 28,390095 | 3,47489 | 8,170 | <0,001 |  |
| Tempo\_site | -0,201301 | 0,02636 | -7,637 | <0,001 |  |
| Idade | 0,176620 | 0,03082 | 5,731 | <0,001 |  |
| Renda\_região | -0,000133 | 0,00002 | -6,246 | <0,001 |  |
| Tempo\_internet | -0,069548 | 0,00894 | -7,776 | <0,001 |  |

Qual é, aproximadamente, a probabilidade média de clicar no anúncio de um usuário com 34 anos, que mora em uma região cuja renda média é de $ 42.760 e gasta, em média, cerca de 2h40 minutos por dia na internet e 1h30 minutos no site da empresa?

Escolha uma:

a. 16%.

b. 63%.

c. 84%.

d. 37%.

**Feedback**

Sua resposta está correta.

A resposta correta é: 37%..

**Questão 5**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Uma empresa de concessão de crédito tem o intuito de avaliar a probabilidade de que seus clientes não cumpram com seus compromissos de pagamentos (probabilidade de default). Por meio de uma base de dados com 2.000 clientes que obtiveram crédito recentemente, a empresa pretende ajustar um modelo de regressão logística binária, utilizando como variáveis explicativas a idade, o sexo (feminino = 0; masculino = 1) e a renda mensal (R$) de cada indivíduo. A variável dependente se refere ao default propriamente dito (não default = 0; default = 1).

Output do modelo de Regressão Logística estimado para o problema de concessão de crédito - covariáveis: idade, sexo e renda.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | 2,9751 | 0,2623 | 11,34 | < 0,001 |  |
| idade | -0,0243 | 0,0070 | -3,49 | < 0,001 |  |
| sexo\_masc | 0,7415 | 0,1135 | 6,53 | < 0,001 |  |
| renda | -0,00026 | 0,0000 | -15,03 | < 0,001 |  |

Número de observações = 2000

Avalie a tabela de resumo de resultados do modelo e escolha a alternativa correta.

Escolha uma:

a. Como o p-valor de todos os testes < 0,05, pode-se afirmar que as estimativas das variáveis explicativas são estatisticamente diferentes de zero, ao nível de significância de 5% e, portanto, as 3 covariáveis devem ser consideradas no modelo final.

b. Como os coeficientes das variáveis idade e renda < 0,05, então apenas essas duas variáveis devem compor o modelo final.

c. Como o p-valor de todos os testes < 0,05, pode-se afirmar que as estimativas das variáveis explicativas são estatisticamente iguais a zero (0), ao nível de significância de 5%.

d. Como o sinal dos coeficientes das variáveis idade e renda são negativos, então essas variáveis não devem compor o modelo.

**Feedback**

Sua resposta está correta.

A resposta correta é: Como o p-valor de todos os testes < 0,05, pode-se afirmar que as estimativas das variáveis explicativas são estatisticamente diferentes de zero, ao nível de significância de 5% e, portanto, as 3 covariáveis devem ser consideradas no modelo final..

**Questão 6**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Considere a área de Digital Analytics de uma empresa varejista que está iniciando sua operação no e-commerce, e deseja  melhorar  a eficácia de sua publicidade a partir de impressões de seus anúncios baseado no perfil do usuário. Para isso, está avaliando diversas informações coletadas a partir da navegação web, e deseja estimar a probabilidade de clique do usuário no anúncio apresentado.

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| Idade | Idade do usuário em anos |  |
|  |
| Renda\_região | Renda média da região na qual o usuário está localizado |  |
|  |
| Gênero | 1, se masculino e 0, se feminino |  |
|  |
| Tempo\_internet | Tempo gasto por dia na internet (em minutos) - obtido a partir dos dados da DMP (Data Management Platform) |  |
|  |
| Tempo\_site | Tempo gasto por dia no site da varejista (em minutos) - obtido com o antigo pixel instalado no site |  |
|  |
| Clicou | Flag se o usuário clicou no anúncio (1) ou não clicou (0) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de clique no anúncio - covariáveis: Idade, Renda\_região, Gênero, Tempo\_internet e Tempo\_site.

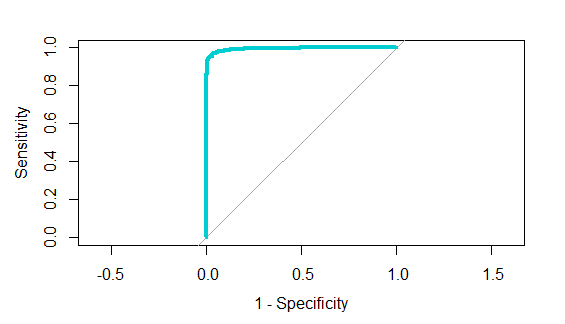
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| (Intercept) | 28,51771 | 3,49556 | 8158,00000 | <0,001 |  |
| Tempo\_site | -0,20172 | 0,02647 | -7,62100 | <0,001 |  |
| Idade | 0,17601 | 0,03093 | 5,69000 | <0,001 |  |
| Randa\_região | -0,00013 | 0,00002 | -6,24500 | <0,001 |  |
| Tempo\_internet | -0,06931 | 0,00896 | -7,73900 | <0,001 |  |
| Gênero | -0,20946 | 0,50116 | -0,41800 | 0,6760 |  |

Número de observações = 700.

Output obtido após o processo de redução de variáveis do modelo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| Intercepto | 28,390095 | 3,47489 | 8,170 | <0,001 |  |
| Tempo\_site | -0,201301 | 0,02636 | -7,637 | <0,001 |  |
| Idade | 0,176620 | 0,03082 | 5,731 | <0,001 |  |
| Randa\_região | -0,000133 | 0,00002 | -6,246 | <0,001 |  |
| Tempo\_internet | -0,069548 | 0,00894 | -7,776 | <0,001 |  |

A análise da área abaixo da curva ROC indicada abaixo nos indica que:



Em que: sensitivity = senbilidade e specificity = especificidade.

Escolha uma:

a. O acerto do modelo parece ruim, já que a curva está bem distante da linha diagonal e esperamos uma métrica AUC muito próxima de 1.

b. O acerto do modelo parece bom, já que a curva está bem distante da linha diagonal e esperamos uma métrica AUC muito próxima de 1.

c. O acerto do modelo parece bom, já que a curva está bem distante da linha diagonal e esperamos uma métrica AUC muito próxima de 0.

d. O acerto do modelo parece ruim, já que a curva está bem distante da linha diagonal e esperamos uma métrica AUC muito próxima de 0.

**Feedback**

Sua resposta está correta.

A resposta correta é: O acerto do modelo parece bom, já que a curva está bem distante da linha diagonal e esperamos uma métrica AUC muito próxima de 1..

**Questão 7**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Considere que o departamento de RH,  por meio da sua área de People Analytics, tenha o interesse em investigar as variáveis explicativas  que influenciam na probabilidade de um trainee chegar atrasado à empresa. Sendo assim, o RH elaborou uma pesquisa com os 100 trainees perguntando se ele costuma ou não chegar atrasado no trabalho. Perguntou também sobre a distância percorrida no trajeto (em quilômetros), o número de semáforos pelos quais cada um passou, o período em que foi realizado o trajeto (manhã ou tarde) e como cada um se considera em termos de perfil ao volante (calmo, moderado ou agressivo).

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| tempo | Tempo para chegar à empresa (em minutos) |  |
|  |
| dist | Distância percorrida até a empresa (quilômetros) |  |
|  |
| sem | Quantidade de semáforos |  |
|  |
| per | Período do dia no qual o trainee foi à empresa. Manhã (categoria de referência - dummy = 1): caso o trainee tenha ido mais cedo para atividades específicas de sua área ou Tarde (dummy = 0): se o trainee foi apenas para o período da tarde, horário definido para as atividades pertinentes ao programa de Trainee |  |
|  |
|  |
| perfil | Perfil ao volante (Perfil 1: calmo, Perfil 2: moderado e Perfil 3: agressivo) |  |
|  |

Indique para área de People Analytics a maneira mais adequada para considerar a informação de perfil ao volante na estimação do modelo.

Escolha uma:

a. Por meio da criação de variáveis dummies, nesse caso, serão necessárias 2 variávies dummies.

b. Criar uma variável numérica que assume o valor 1 para o perfil calmo, o valor 2 para o perfil moderado e o valor 3 para o perfil agressivo.

c. Por meio da criação de variáveis dummies, nesse caso, serão necessárias 3 variávies dummies.

d. Por meio da criação de variáveis dummies, nesse caso, será necessária 1 variáviel dummy.

**Feedback**

Sua resposta está correta.

A resposta correta é: Por meio da criação de variáveis dummies, nesse caso, serão necessárias 2 variávies dummies..

**Questão 8**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Considere a área de Digital Analytics de uma empresa varejista que está iniciando sua operação no e-commerce, e deseja  melhorar  a eficácia de sua publicidade a partir de impressões de seus anúncios baseado no perfil do usuário. Para isso, está avaliando diversas informações coletadas a partir da navegação web, e deseja estimar a probabilidade de clique do usuário no anúncio apresentado.

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| Idade | Idade do usuário em anos |  |
|  |
| Renda\_região | Renda média da região na qual o usuário está localizado |  |
|  |
| Gênero | 1, se masculino e 0, se feminino |  |
|  |
| Tempo\_internet | Tempo gasto por dia na internet (em minutos) - obtido a partir dos dados da DMP (Data Management Platform) |  |
|  |
| Tempo\_site | Tempo gasto por dia no site da varejista (em minutos) - obtido com o antigo pixel instalado no site |  |
|  |
| Clicou | Flag se o usuário clicou no anúncio (1) ou não clicou (0) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de clique no anúncio - covariáveis: Idade, Renda\_região, Gênero, Tempo\_internet e Tempo\_site.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| (Intercept) | 28,51771 | 3,49556 | 8158,00000 | <0,001 |  |
| Tempo\_site | -0,20172 | 0,02647 | -7,62100 | <0,001 |  |
| Idade | 0,17601 | 0,03093 | 5,69000 | <0,001 |  |
| Randa\_região | -0,00013 | 0,00002 | -6,24500 | <0,001 |  |
| Tempo\_internet | -0,06931 | 0,00896 | -7,73900 | <0,001 |  |
| Gênero | -0,20946 | 0,50116 | -0,41800 | 0,6760 |  |

Número de observações = 700.

Output obtido após o processo de redução de variáveis do modelo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| Intercepto | 28,390095 | 3,47489 | 8,170 | <0,001 |  |
| Tempo\_site | -0,201301 | 0,02636 | -7,637 | <0,001 |  |
| Idade | 0,176620 | 0,03082 | 5,731 | <0,001 |  |
| Randa\_região | -0,000133 | 0,00002 | -6,246 | <0,001 |  |
| Tempo\_internet | -0,069548 | 0,00894 | -7,776 | <0,001 |  |

Considere a análise de desempenho apresentada a seguir e indique a alternativa correta com relação ao desempenho do modelo estimado:

|  |  |  |  |
| --- | --- | --- | --- |
| Observado | Estimado | |  |
| Não clique | Clique | Total |
| Não clique | 151 | 0 | 151 |
| Clique | 7 | 142 | 149 |
|  | 158 | 142 | 300 |

|  |  |
| --- | --- |
| Sensibilidade | 95,30% |
| Especificidade | 100,00% |

Escolha uma:

a. A proporção de casos de não clique corretamente classificados pelo modelo é 100% e o acerto global do modelo é 96,61%.

b. A proporção de casos de não clique corretamente classificados pelo modelo é 97,66% e o acerto global do modelo é 95,30%.

c. A proporção de casos de clique corretamente classificados pelo modelo é 100% e o acerto global do modelo é 97,66%.

d. A proporção de casos de clique corretamente classificados pelo modelo é 95,30% e o acerto global do modelo é 97,66%.

**Feedback**

Sua resposta está correta.

A resposta correta é: A proporção de casos de clique corretamente classificados pelo modelo é 95,30% e o acerto global do modelo é 97,66%..

**Questão 9**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Qual(is) recurso(s) estatístico(s) indicado(s) abaixo é/são comumente utilizado(s) para avaliar o acerto de um modelo de regressão logística?

Escolha uma:

a. Cálculo da correlação entre os resíduos e a variável resposta observada e entre as variáveis explicativas e a variável resposta.

b. Métrica KS (Kolmogorov-Smirnov) e gráfico da Curva ROC.

c. Teste de Normalidade dos resíduos.

d. Gráfico de Dispersão.

**Feedback**

Sua resposta está correta.

A resposta correta é: Métrica KS (Kolmogorov-Smirnov) e gráfico da Curva ROC..

**Questão 10**

Incorreto

Atingiu 0,00 de 1,00

Marcar questão

**Texto da questão**

Considere que o departamento de RH,  por meio da sua área de People Analytics, tenha o interesse em investigar as variáveis explicativas  que influenciam na probabilidade de um trainee chegar atrasado à empresa. Sendo assim, o RH elaborou uma pesquisa com os 100 trainees perguntando se ele costuma ou não chegar atrasado no trabalho. Perguntou também sobre a distância percorrida no trajeto (em quilômetros), o número de semáforos pelos quais cada um passou, o período em que foi realizado o trajeto (manhã ou tarde) e como cada um se considera em termos de perfil ao volante (calmo, moderado ou agressivo).

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| tempo | Tempo para chegar à empresa (em minutos) |  |
|  |
| dist | Distância percorrida até a empresa (quilômetros) |  |
|  |
| sem | Quantidade de semáforos |  |
|  |
| per | Período do dia no qual o trainee foi à empresa. Manhã (categoria de referência - dummy = 1): caso o trainee tenha ido mais cedo para atividades específicas de sua área ou Tarde (dummy = 0): se o trainee foi apenas para o período da tarde, horário definido para as atividades pertinentes ao programa de Trainee |  |
|  |
|  |
| perfil | Perfil ao volante (Perfil 1: calmo, Perfil 2: moderado e Perfil 3: agressivo) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de atraso do trainee - covariáveis: dist, sem, per e perfil.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | -30,2003 | 9,9811 | -3,03 | 0,002 |  |
| dist | 0,2202 | 0,1097 | 2,01 | 0,045 |  |
| sem | 2,7667 | 0,9217 | 3,00 | 0,003 |  |
| per | -3,6534 | 0,8781 | -4,16 | <0,001 |  |
| perfil\_2 | 1,3460 | 0,7477 | 1,80 | 0,072 |  |
| perfil\_3 | 2,91447 | 1,1788 | 2,47 | 0,013 |  |

Número de observações = 100

Output obtido via stepwise do modelo de Regressão Logística estimado para o problema de atraso do trainee - covariáveis: dist, sem, per e perfil.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | -30,9334 | 10,6363 | -2,91 | 0,004 |  |
| dist | 0,2041 | 0,1012 | 2,02 | 0,044 |  |
| sem | 2,9201 | 1,0108 | 2,89 | 0,004 |  |
| per | -3,7763 | 0,8467 | -4,46 | <0,001 |  |
| perfil\_3 | 2,45907 | 1,1395 | 2,16 | 0,031 |  |

Considere a análise de desempenho apresentada a seguir e indique a alternativa correta com relação ao desempenho do modelo estimado:

|  |  |  |  |
| --- | --- | --- | --- |
| Observado | Estimado | | Total |
| Não atraso | Atraso |
| Não atraso | 28 | 13 | 41 |
| Atraso | 2 | 57 | 59 |
| Total | 30 | 70 | 100 |

|  |  |
| --- | --- |
| Sensibilidade | 96,61% |
| Especificidade | 68,29% |

Escolha uma:

a. A proporção de casos de atraso corretamente classificados pelo modelo é 96,61% e o acerto global do modelo é 85,00%.

b. A proporção de casos de atraso corretamente classificados pelo modelo é 68,29% e o acerto global do modelo é 96,61%.

c. A proporção de casos de não atraso corretamente classificados pelo modelo é 96,61% e o acerto global do modelo é 85,00%.

d. A proporção de casos de não atraso corretamente classificados pelo modelo é 68,29% e o acerto global do modelo é 96,61%.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: A proporção de casos de atraso corretamente classificados pelo modelo é 96,61% e o acerto global do modelo é 85,00%..

**Questão 11**

Correto

Atingiu 1,00 de 1,00

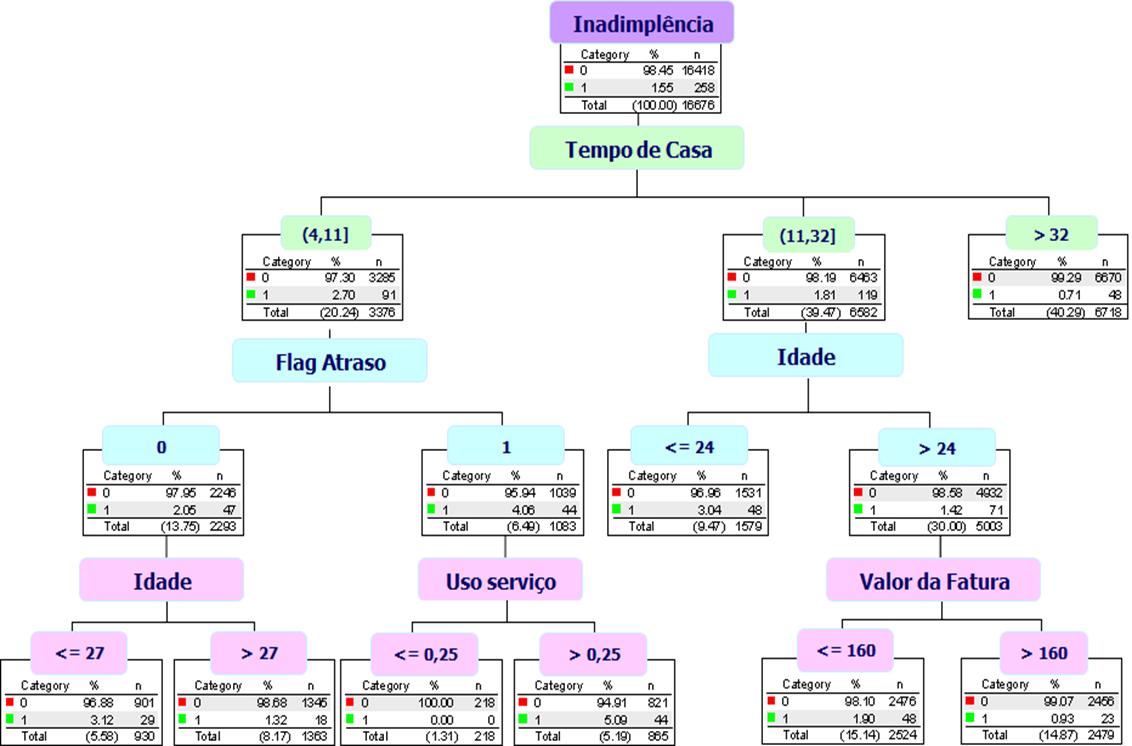
Marcar questão

**Texto da questão**

**Uma fintech com atuação exclusiva no estado do Espírito Santo deseja avaliar a inadimplência esperada de seus clientes no cenário de crise/pós-crise da pandemia de COVID-19. Para isso, reuniu as informações dos últimos 6 meses de seus clientes (Março 2020 a Agosto de 2020) e, com a ajuda de uma consultoria estatística, desenvolveu um modelo de Árvore de Decisão utilizando as seguintes informações:**

* **Inadimplência: atraso ou não pagamento de, pelo menos, uma fatura do cartão no horizonte de 3 meses (inadimplência = 1, e adimplência = 0);**
* **Tempo de Casa: tempo de relacionamento entre o cliente e a fintech (em meses);**
* **Idade: idade do cliente (em anos);**
* **Flag Atraso: flag se o cliente possui atraso no mercado, oriundo de um Bureau de Crédito;**
* **Uso serviço: percentual de vezes em que o cliente utilizou milhas para pagamento de contas no aplicativo, nos últimos 3 meses;**
* **Valor da Fatura: valor médio das últimas 3 faturas do cliente (em R$).**

**Em cada nó da Árvore de Decisão, no diagrama a seguir, o par de valores abaixo de "%" representam o percentual de clientes adimplentes (0) e inadimplentes (1) dentro do respectivo nó. Os dois valores abaixo de "n" representam as quantidades absolutas de clientes adimplentes e inadimplentes, respectivamente. Já os números da última linha, dentro de cada nó, correspondem à sua representatividade na base de dados, em percentual e em número absoluto.**



A interpretação do nó que agrupa os clientes com tempo de casa entre (11,32] meses e Idade <= 24 nos indica o quê?

Escolha uma:

a. Há 1.531 clientes com o perfil de interesse, representando 9,47% do total da base e a taxa de inadimplência desse público é 3,04%.

b. Há 1.579 clientes com o perfil de interesse, representando 3,04% do total da base e a taxa de inadimplência desse público é 9,47%.

c. Há 1.579 clientes com o perfil de interesse, representando 9,47% do total da base e a taxa de inadimplência desse público é 3,04%.

d. Há 1.579 clientes com o perfil de interesse, cuja taxa de inadimplência desse público é 9,47%.

**Feedback**

Sua resposta está correta.

A resposta correta é: Há 1.579 clientes com o perfil de interesse, representando 9,47% do total da base e a taxa de inadimplência desse público é 3,04%..

**Questão 12**

Correto

Atingiu 1,00 de 1,00

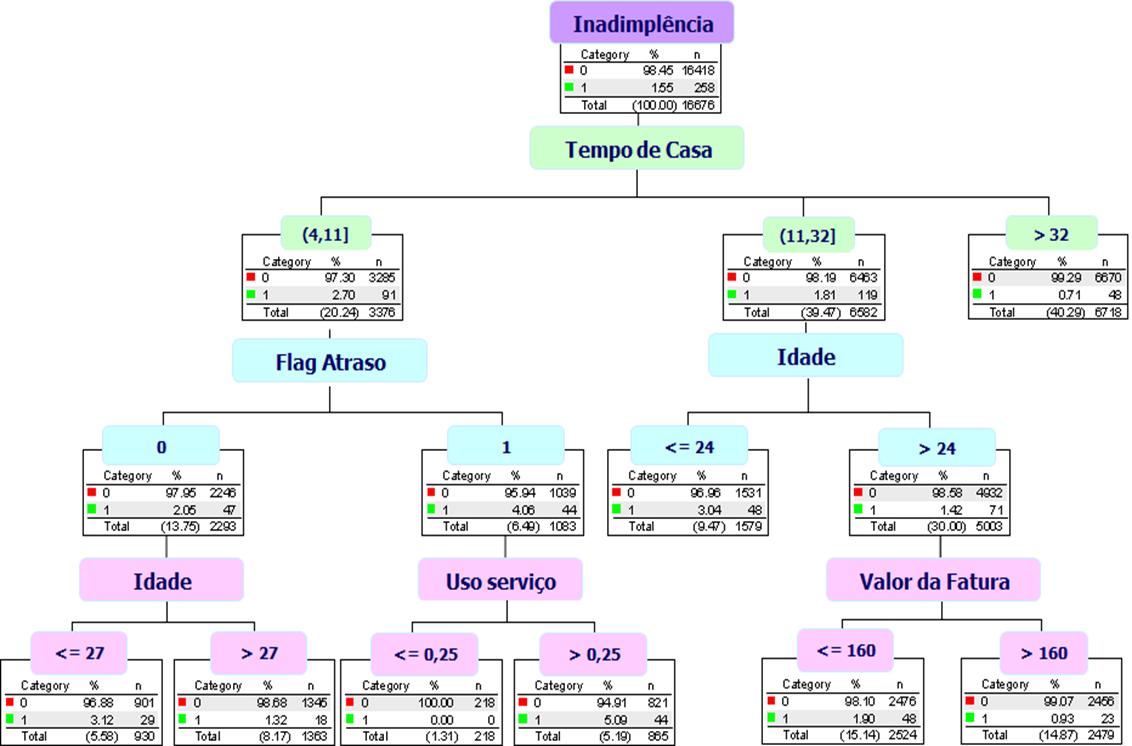
Marcar questão

**Texto da questão**

**Uma fintech com atuação exclusiva no estado do Espírito Santo deseja avaliar a inadimplência esperada de seus clientes no cenário de crise/pós-crise da pandemia de COVID-19. Para isso, reuniu as informações dos últimos 6 meses de seus clientes (Março 2020 a Agosto de 2020) e, com a ajuda de uma consultoria estatística, desenvolveu um modelo de Árvore de Decisão utilizando as seguintes informações:**

* **Inadimplência: atraso ou não pagamento de, pelo menos, uma fatura do cartão no horizonte de 3 meses (inadimplência = 1, e adimplência = 0);**
* **Tempo de Casa: tempo de relacionamento entre o cliente e a fintech (em meses);**
* **Idade: idade do cliente (em anos);**
* **Flag Atraso: flag se o cliente possui atraso no mercado, oriundo de um Bureau de Crédito;**
* **Uso serviço: percentual de vezes em que o cliente utilizou milhas para pagamento de contas no aplicativo, nos últimos 3 meses;**
* **Valor da Fatura: valor médio das últimas 3 faturas do cliente (em R$).**

**Em cada nó da Árvore de Decisão, no diagrama a seguir, o par de valores abaixo de "%" representam o percentual de clientes adimplentes (0) e inadimplentes (1) dentro do respectivo nó. Os dois valores abaixo de "n" representam as quantidades absolutas de clientes adimplentes e inadimplentes, respectivamente. Já os números da última linha, dentro de cada nó, correspondem à sua representatividade na base de dados, em percentual e em número absoluto.**



A interpretação do nó que agrupa os clientes com tempo de casa > 32 meses nos indica o quê?

Escolha uma:

a. Há 6.718 clientes com o perfil de interesse, representando 0,71% do total da base e a taxa de inadimplência desse público é 40,29%.

b. Há 6.718 clientes com o perfil de interesse, representando 40,29% do total da base e a taxa de inadimplência desse público é 0,71%.

c. Há 1.579 clientes com o perfil de interesse, representando 3,04% do total da base.

d. É o nó com a maior taxa de inadimplência: 67,18%.

**Feedback**

Sua resposta está correta.

A resposta correta é: Há 6.718 clientes com o perfil de interesse, representando 40,29% do total da base e a taxa de inadimplência desse público é 0,71%..

**Questão 13**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

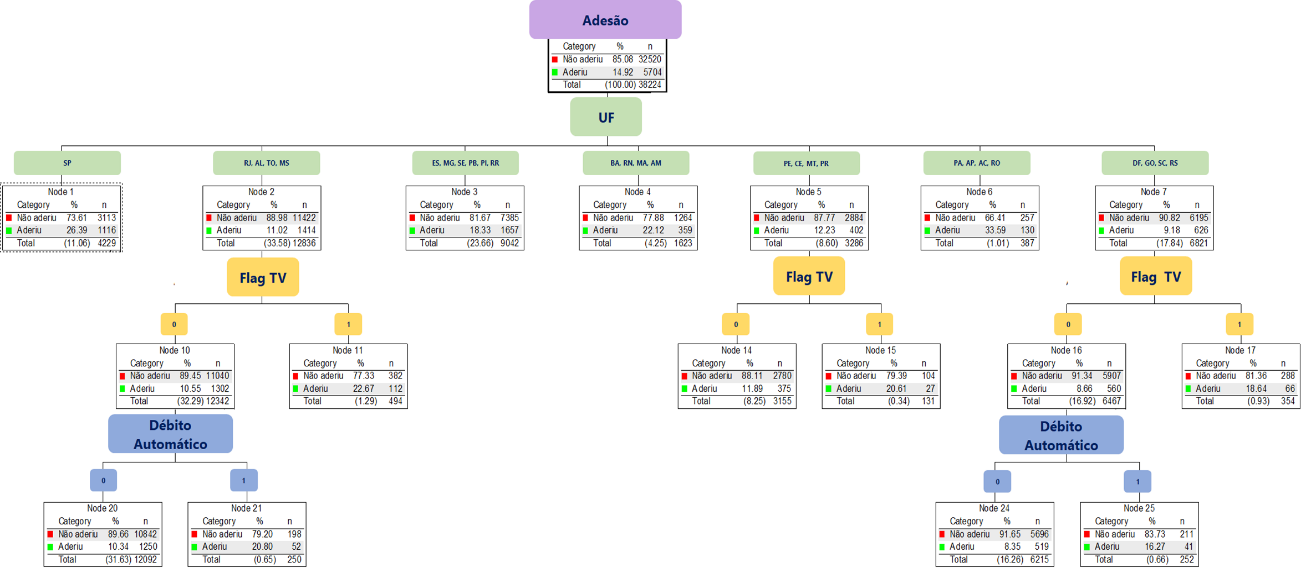
Uma empresa de telecomunicações gostaria de avaliar o perfil dos clientes que aderiram, nos três primeiros meses de lançamento, à utilização do seu novo aplicativo que permite assistir determinados conteúdos televisivos no celular (ou por meio de compartilhamento entre celular e televisões smart) por apenas R$ 5,99 adicionais no valor do pacote já contratado. Para avaliação da propensão, foi utilizado o algoritmo o de Árvore de Decisão com uma amostra aleatória de aproximadamente 38 mil clientes, e foram consideradas as seguintes informações:

Adesão: adesão ao aplicativo lançado (1: aderiu, 0: não aderiu).

UF: Estado de residência do cliente.

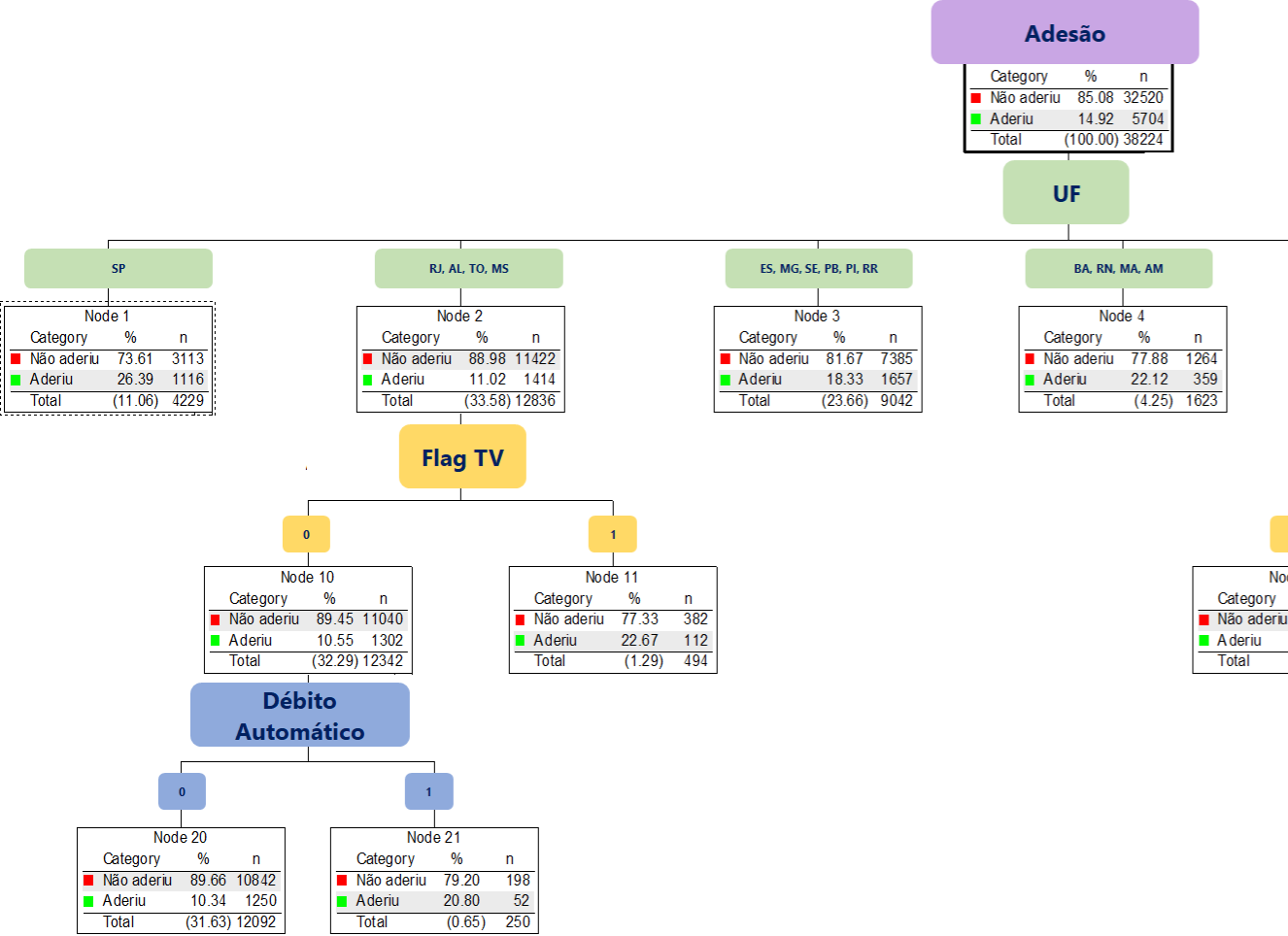
Flag TV: indicativo se o cliente possui o produto de canais de TV Paga no pacote.

*Débito automático:* flag se o cliente paga por débito automático.

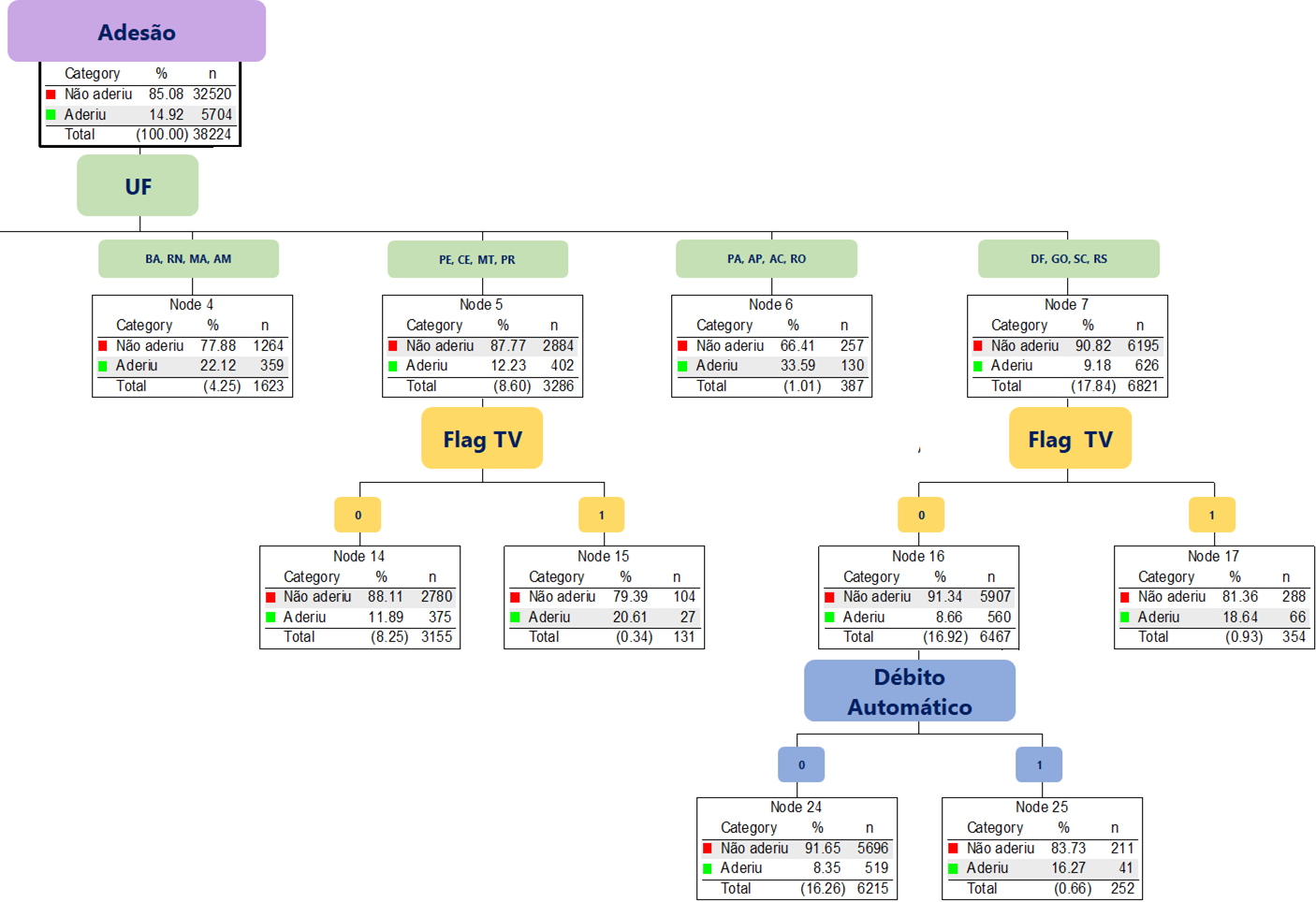


Visualizar árvore mais de perto por partes:

**Lado ESQUERDO**



**Lado DIREITO**



Quantas variáveis entraram na árvore?

Escolha uma:

a. 4 variáveis.

b. 3 variáveis.

c. 5 variáveis.

d. 12 variáveis.

**Feedback**

Sua resposta está correta.

A resposta correta é: 3 variáveis..

**Questão 14**

Incorreto

Atingiu 0,00 de 1,00

Marcar questão

**Texto da questão**

Sobre o algoritmo utilizado na Árvore de Decisão, indique a alternativa incorreta:

Escolha uma:

a. Os nós finais podem ser classificados em propensos ou não propensos ao evento de interesse utilizando como ponto de corte a proporção do evento resposta da base inicial.

b. Podem ser testadas variáveis quantitativas apenas.

c. Uma mesma variável pode ser utilizada em mais de um nível da árvore.

d. O algoritmo seleciona dentre todas as variáveis explicativas disponíveis, aquela que melhor discrimina a variável resposta, segundo um critério de partição, por exemplo, o Teste Qui-quadrado.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: Podem ser testadas variáveis quantitativas apenas..

**Questão 15**

Incorreto

Atingiu 0,00 de 1,00

Marcar questão

**Texto da questão**

Sobre o teste Qui-Quadrado podemos afirmar que:

Escolha uma:

a. O objetivo do teste do Qui-Quadrado é verificar se existe associação entre duas variáveis qualitativas.

b. A hipótese nula do teste é a de que existe associação entre as variáveis testadas.

c. A hipótese nula do teste é a de que existe correlação linear positiva entre as variáveis testadas.

d. O objetivo do teste do Qui-Quadrado é verificar se existe associação entre duas variáveis quantitativas.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: O objetivo do teste do Qui-Quadrado é verificar se existe associação entre duas variáveis qualitativas..

**Questão 16**

Correto

Atingiu 1,00 de 1,00

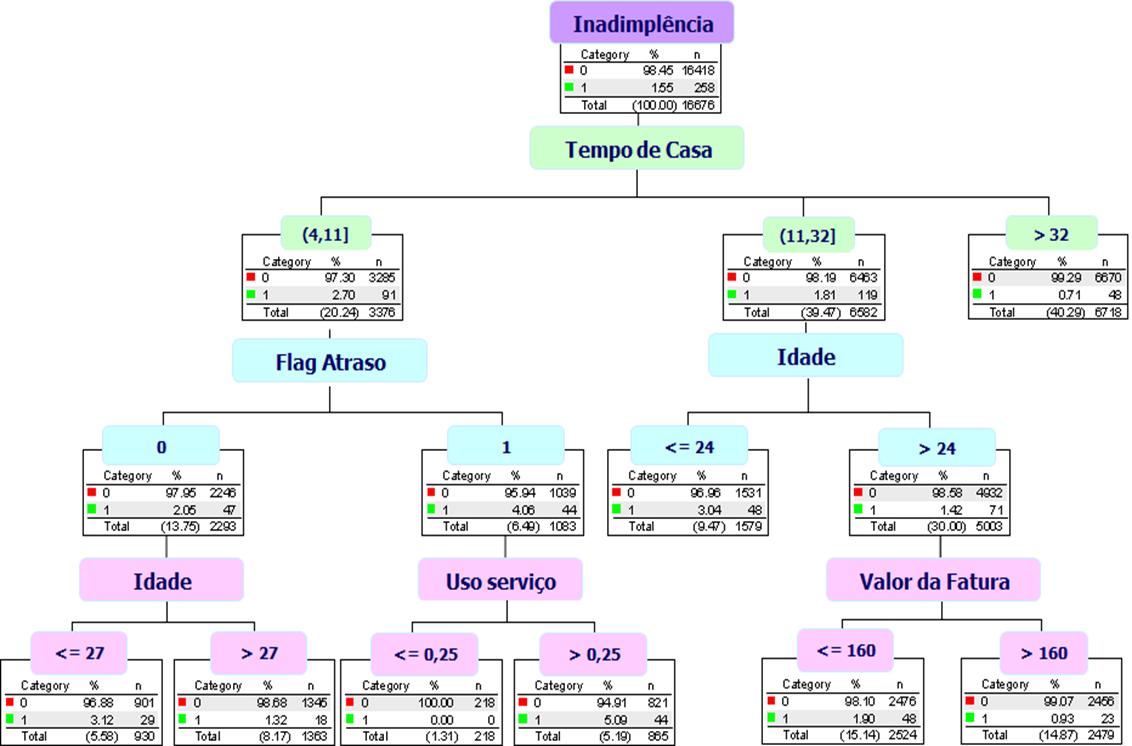
Marcar questão

**Texto da questão**

**Uma fintech com atuação exclusiva no estado do Espírito Santo deseja avaliar a inadimplência esperada de seus clientes no cenário de crise/pós-crise da pandemia de COVID-19. Para isso, reuniu as informações dos últimos 6 meses de seus clientes (Março 2020 a Agosto de 2020) e, com a ajuda de uma consultoria estatística, desenvolveu um modelo de Árvore de Decisão utilizando as seguintes informações:**

* **Inadimplência: atraso ou não pagamento de, pelo menos, uma fatura do cartão no horizonte de 3 meses (inadimplência = 1, e adimplência = 0);**
* **Tempo de Casa: tempo de relacionamento entre o cliente e a fintech (em meses);**
* **Idade: idade do cliente (em anos);**
* **Flag Atraso: flag se o cliente possui atraso no mercado, oriundo de um Bureau de Crédito;**
* **Uso serviço: percentual de vezes em que o cliente utilizou milhas para pagamento de contas no aplicativo, nos últimos 3 meses;**
* **Valor da Fatura: valor médio das últimas 3 faturas do cliente (em R$).**

**Em cada nó da Árvore de Decisão, no diagrama a seguir, o par de valores abaixo de "%" representam o percentual de clientes adimplentes (0) e inadimplentes (1) dentro do respectivo nó. Os dois valores abaixo de "n" representam as quantidades absolutas de clientes adimplentes e inadimplentes, respectivamente. Já os números da última linha, dentro de cada nó, correspondem à sua representatividade na base de dados, em percentual e em número absoluto.**



Quantos nós finais (perfis) existem na árvore?

Escolha uma:

a. 13 nós finais.

b. 8 nós finais.

c. 6 nós finais.

d. 14 nós finais.

**Feedback**

Sua resposta está correta.

A resposta correta é: 8 nós finais..

**Questão 17**

Correto

Atingiu 1,00 de 1,00

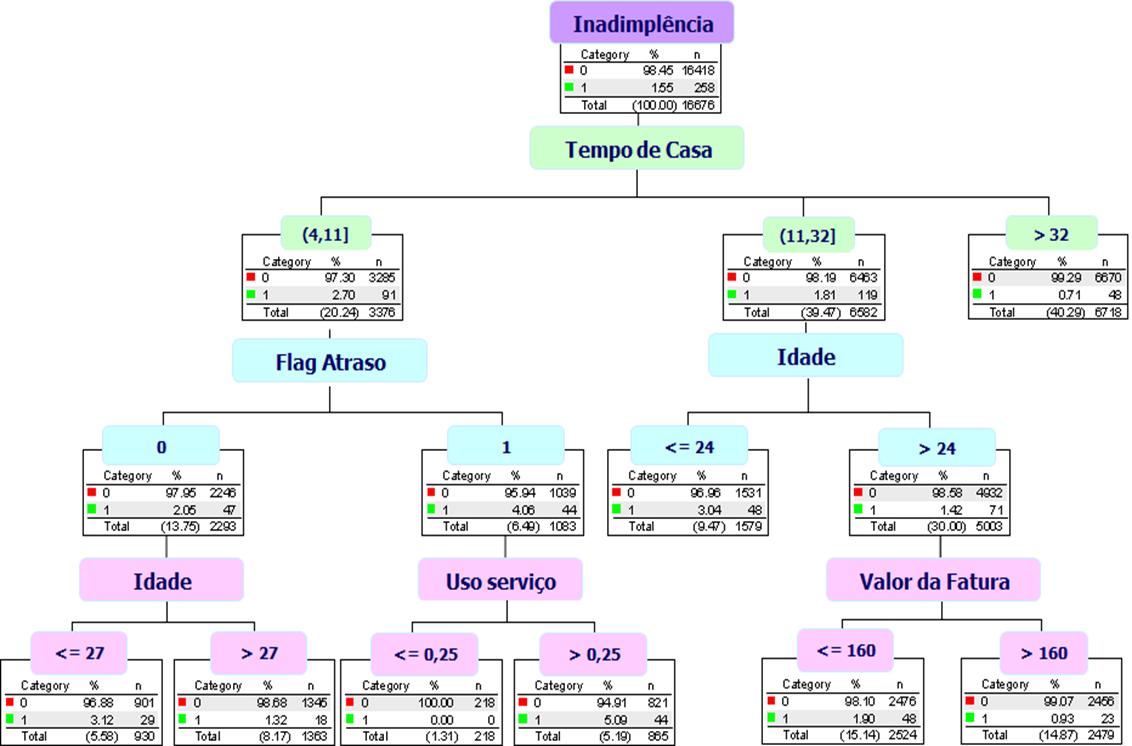
Marcar questão

**Texto da questão**

**Uma fintech com atuação exclusiva no estado do Espírito Santo deseja avaliar a inadimplência esperada de seus clientes no cenário de crise/pós-crise da pandemia de COVID-19. Para isso, reuniu as informações dos últimos 6 meses de seus clientes (Março 2020 a Agosto de 2020) e, com a ajuda de uma consultoria estatística, desenvolveu um modelo de Árvore de Decisão utilizando as seguintes informações:**

* **Inadimplência: atraso ou não pagamento de, pelo menos, uma fatura do cartão no horizonte de 3 meses (inadimplência = 1, e adimplência = 0);**
* **Tempo de Casa: tempo de relacionamento entre o cliente e a fintech (em meses);**
* **Idade: idade do cliente (em anos);**
* **Flag Atraso: flag se o cliente possui atraso no mercado, oriundo de um Bureau de Crédito;**
* **Uso serviço: percentual de vezes em que o cliente utilizou milhas para pagamento de contas no aplicativo, nos últimos 3 meses;**
* **Valor da Fatura: valor médio das últimas 3 faturas do cliente (em R$).**

**Em cada nó da Árvore de Decisão, no diagrama a seguir, o par de valores abaixo de "%" representam o percentual de clientes adimplentes (0) e inadimplentes (1) dentro do respectivo nó. Os dois valores abaixo de "n" representam as quantidades absolutas de clientes adimplentes e inadimplentes, respectivamente. Já os números da última linha, dentro de cada nó, correspondem à sua representatividade na base de dados, em percentual e em número absoluto.**



Quantos níveis de variáveis explicativas existem na árvore?

Escolha uma:

a. 4 níveis.

b. 6 níveis.

c. 3 níveis.

d. 5 níveis.

**Feedback**

Sua resposta está correta.

A resposta correta é: 3 níveis..

**Questão 18**

Correto

Atingiu 1,00 de 1,00

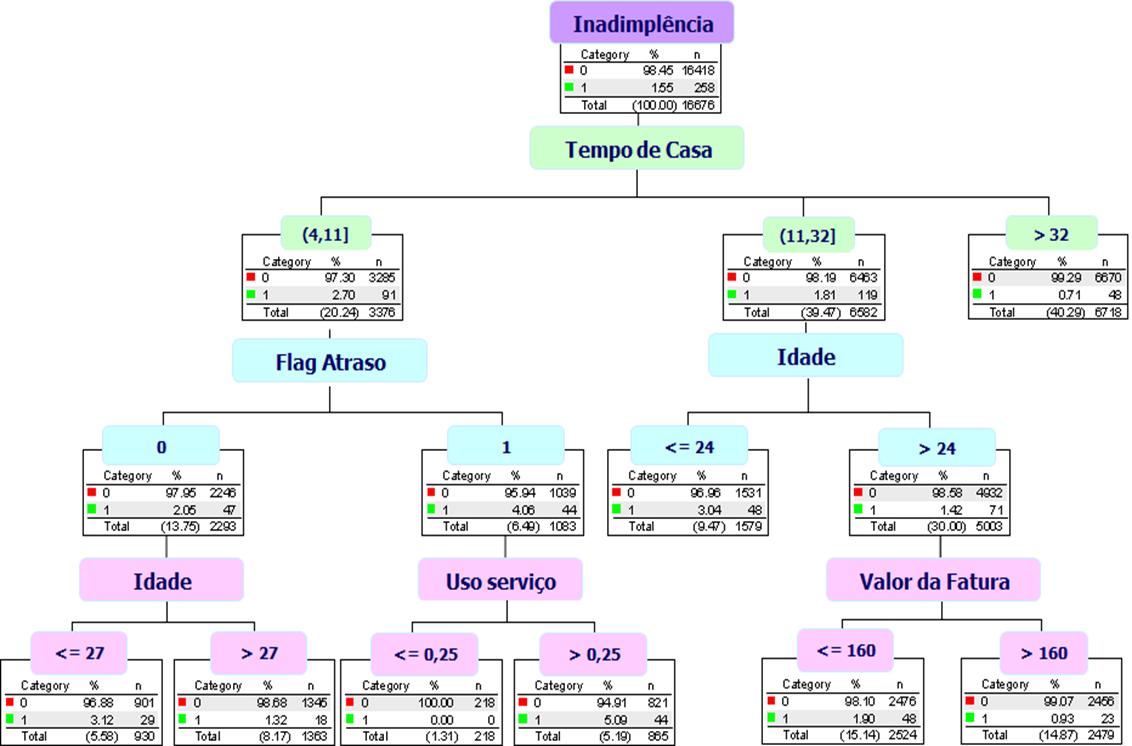
Marcar questão

**Texto da questão**

**Uma fintech com atuação exclusiva no estado do Espírito Santo deseja avaliar a inadimplência esperada de seus clientes no cenário de crise/pós-crise da pandemia de COVID-19. Para isso, reuniu as informações dos últimos 6 meses de seus clientes (Março 2020 a Agosto de 2020) e, com a ajuda de uma consultoria estatística, desenvolveu um modelo de Árvore de Decisão utilizando as seguintes informações:**

* **Inadimplência: atraso ou não pagamento de, pelo menos, uma fatura do cartão no horizonte de 3 meses (inadimplência = 1, e adimplência = 0);**
* **Tempo de Casa: tempo de relacionamento entre o cliente e a fintech (em meses);**
* **Idade: idade do cliente (em anos);**
* **Flag Atraso: flag se o cliente possui atraso no mercado, oriundo de um Bureau de Crédito;**
* **Uso serviço: percentual de vezes em que o cliente utilizou milhas para pagamento de contas no aplicativo, nos últimos 3 meses;**
* **Valor da Fatura: valor médio das últimas 3 faturas do cliente (em R$).**

**Em cada nó da Árvore de Decisão, no diagrama a seguir, o par de valores abaixo de "%" representam o percentual de clientes adimplentes (0) e inadimplentes (1) dentro do respectivo nó. Os dois valores abaixo de "n" representam as quantidades absolutas de clientes adimplentes e inadimplentes, respectivamente. Já os números da última linha, dentro de cada nó, correspondem à sua representatividade na base de dados, em percentual e em número absoluto.**



Quantas variáveis explicativas compuseram a árvore?

Escolha uma:

a. 5 variáveis.

b. 8 variáveis.

c. 7 variáveis.

d. 6 variáveis.

**Feedback**

Sua resposta está correta.

A resposta correta é: 5 variáveis..

**Questão 19**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Uma empresa de concessão de crédito tem o intuito de avaliar a probabilidade de que seus clientes não cumpram com seus compromissos de pagamentos (probabilidade de default). Por meio de uma base de dados com 2.000 clientes que obtiveram crédito recentemente, a empresa pretende ajustar um modelo de regressão logística binária, utilizando como variáveis explicativas a idade, o sexo (feminino = 0; masculino = 1) e a renda mensal (R$) de cada indivíduo. A variável dependente se refere ao default propriamente dito (não default = 0; default = 1).

Output do modelo de Regressão Logística estimado para o problema de concessão de crédito - covariáveis: idade, sexo e renda.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | 2,9751 | 0,2623 | 11,34 | < 0,001 |  |
| idade | -0,0243 | 0,0070 | -3,49 | < 0,001 |  |
| sexo\_masc | 0,7415 | 0,1135 | 6,53 | < 0,001 |  |
| renda | -0,00026 | 0,0000 | -15,03 | < 0,001 |  |

Número de observações = 2000

Com relação à interpretação da variável sexo, escolha a alternativa correta:

Escolha uma:

a. Como o sinal do parâmetro estimado para a variável sexo é positivo, os indivíduos do sexo masculino (dummy = 1) apresentam, em média, maiores probabilidades de default do que os do sexo feminino, mantidas as demais condições constantes.

b. Como o sinal do parâmetro estimado para a variável sexo é maior do que 0, os indivíduos do sexo masculino (dummy = 1) apresentam, em média, menores probabilidades de default do que os do sexo feminino, mantidas as demais condições constantes.

c. O sexo feminino apresenta uma probabilidade de default equivalente a 0,7415.

d. Como o parâmetro estimado para a variável sexo é menor do que 1, os indivíduos do sexo masculino (dummy = 1) apresentam, em média, menores probabilidades de default do que os do sexo feminino, mantidas as demais condições constantes.

**Feedback**

Sua resposta está correta.

A resposta correta é: Como o sinal do parâmetro estimado para a variável sexo é positivo, os indivíduos do sexo masculino (dummy = 1) apresentam, em média, maiores probabilidades de default do que os do sexo feminino, mantidas as demais condições constantes..

**Questão 20**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Considere a área de Digital Analytics de uma empresa varejista que está iniciando sua operação no e-commerce, e deseja  melhorar  a eficácia de sua publicidade a partir de impressões de seus anúncios baseado no perfil do usuário. Para isso, está avaliando diversas informações coletadas a partir da navegação web, e deseja estimar a probabilidade de clique do usuário no anúncio apresentado.

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| Idade | Idade do usuário em anos |  |
|  |
| Renda\_região | Renda média da região na qual o usuário está localizado |  |
|  |
| Gênero | 1, se masculino e 0, se feminino |  |
|  |
| Tempo\_internet | Tempo gasto por dia na internet (em minutos) - obtido a partir dos dados da DMP (Data Management Platform) |  |
|  |
| Tempo\_site | Tempo gasto por dia no site da varejista (em minutos) - obtido com o antigo pixel instalado no site |  |
|  |
| Clicou | Flag se o usuário clicou no anúncio (1) ou não clicou (0) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de clique no anúncio - covariáveis: Idade, Renda\_região, Gênero, Tempo\_internet e Tempo\_site.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| (Intercept) | 28,51771 | 3,49556 | 8158,00000 | <0,001 |  |
| Tempo\_site | -0,20172 | 0,02647 | -7,62100 | <0,001 |  |
| Idade | 0,17601 | 0,03093 | 5,69000 | <0,001 |  |
| Renda\_região | -0,00013 | 0,00002 | -6,24500 | <0,001 |  |
| Tempo\_internet | -0,06931 | 0,00896 | -7,73900 | <0,001 |  |
| Gênero | -0,20946 | 0,50116 | -0,41800 | 0,6760 |  |

Número de observações = 700.

Output obtido após o processo de redução de variáveis do modelo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| Intercepto | 28,390095 | 3,47489 | 8,170 | <0,001 |  |
| Tempo\_site | -0,201301 | 0,02636 | -7,637 | <0,001 |  |
| Idade | 0,176620 | 0,03082 | 5,731 | <0,001 |  |
| Renda\_região | -0,000133 | 0,00002 | -6,246 | <0,001 |  |
| Tempo\_internet | -0,069548 | 0,00894 | -7,776 | <0,001 |  |

Qual das equações abaixo representa adequadamento o cálculo da probabilidade média de clicar no anúncio?

Escolha uma:

a. 

b. 

c. 

d. 

**Feedback**

Sua resposta está correta.

A resposta correta é: .

**Questão 21**

Correto

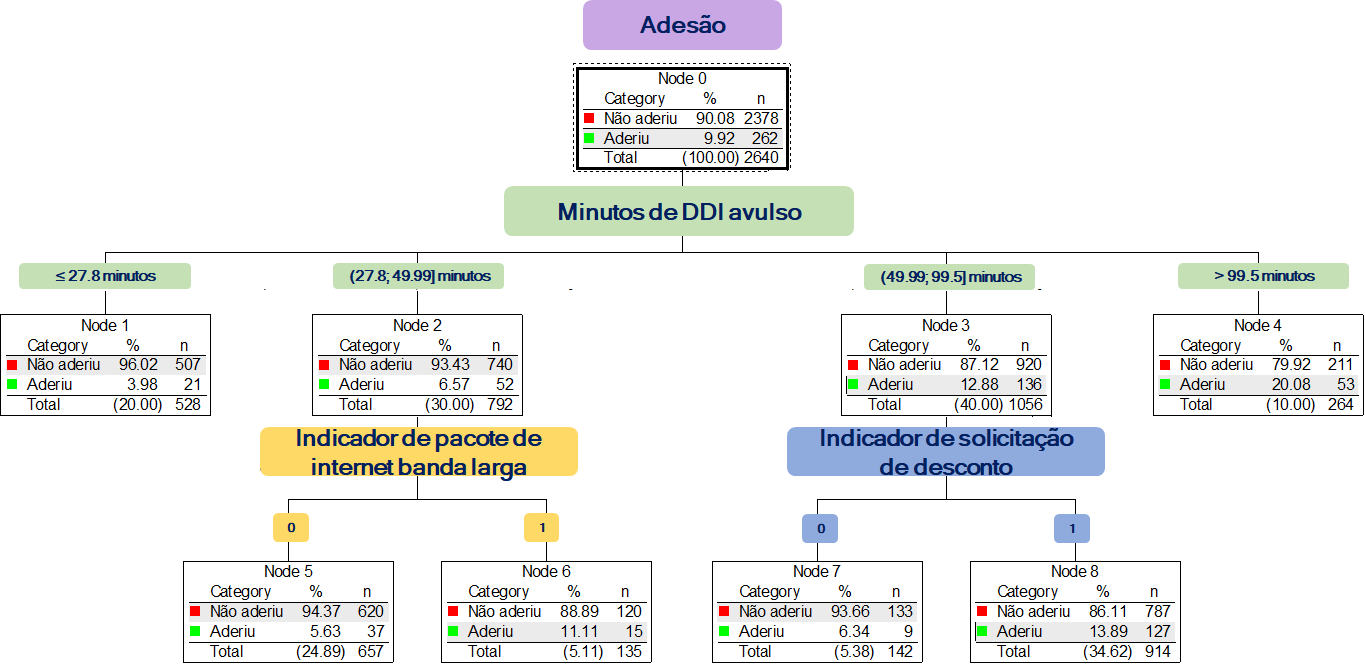
Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Uma companhia telefônica identificou em sua base de clientes um perfil de empresas de pequeno porte que, comparado com a utilização histórica, estavam utilizando com menor frequência o serviço de chamadas internacionais. O analista de negócios responsável pela análise atribuiu essa queda de consumo aos novos softwares de conferência remota, amplamente divulgados nos últimos meses. Sendo assim, com o intuito de reter essa demanda, ofereceu a essas empresas um pacote mensal de DDI com melhores condições de contratação. A árvore abaixo representa a adesão a esse pacote no primeiro mês de oferta:

* Minutos de DDI avulso: média de minutos por dia em chamadas DDI nos últimos 3 meses.
* Indicador de pacote de internet banda larga: indicador de posse de pacote de internet banda larga na empresa.
* Indicador de solicitação de desconto: indicador de solicitação de desconto pelo cliente nos últimos 3 meses em algum dos serviços contratados pelo cliente.



A interpretação do nó que agrupa as empresas que utilizaram entre (27,8; 49,99] minutos por dia em chamadas DDI nos últimos 3 meses nos indica o quê?

Escolha uma:

a. Há 792 empresas com o perfil de interesse, representando 30% do total da base e a taxa de adesão desse público é 6,57%.

b. Há 740 empresas com o perfil de interesse, representando 93,43% do total da base e a taxa de adesão desse público é 30%.

c. Há 52 empresas com o perfil de interesse, representando 6,57% do total da base e a taxa de adesão desse público é 30%.

d. Há 792 empresas com o perfil de interesse, representando 6,57% do total da base e a taxa de adesão desse público é 30%.

**Feedback**

Sua resposta está correta.

A resposta correta é: Há 792 empresas com o perfil de interesse, representando 30% do total da base e a taxa de adesão desse público é 6,57%..

**Questão 22**

Incorreto

Atingiu 0,00 de 1,00

Marcar questão

**Texto da questão**

Uma empresa de concessão de crédito tem o intuito de avaliar a probabilidade de que seus clientes não cumpram com seus compromissos de pagamentos (probabilidade de default). Por meio de uma base de dados com 2.000 clientes que obtiveram crédito recentemente, a empresa pretende ajustar um modelo de regressão logística binária, utilizando como variáveis explicativas a idade, o sexo (feminino = 0; masculino = 1) e a renda mensal (R$) de cada indivíduo. A variável dependente se refere ao default propriamente dito (não default = 0; default = 1).

Output do modelo de Regressão Logística estimado para o problema de concessão de crédito - covariáveis: idade, sexo e renda.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | 2,9751 | 0,2623 | 11,34 | < 0,001 |  |
| idade | -0,0243 | 0,0070 | -3,49 | < 0,001 |  |
| sexo\_masc | 0,7415 | 0,1135 | 6,53 | < 0,001 |  |
| renda | -0,00026 | 0,0000 | -15,03 | < 0,001 |  |

Número de observações = 2000

Considerando as covariáveis presentes no modelo, qual o perfil menos propenso a apresentar default?

Escolha uma:

a. Homens mais jovens, independentemente do rendimento salarial.

b. Homens mais jovens com maiores rendimentos salariais.

c. Mulheres mais jovens, independentemente do rendimento salarial.

d. Mulheres mais velhas com maiores rendimentos salariais.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: Mulheres mais velhas com maiores rendimentos salariais..

**Questão 23**

Correto

Atingiu 1,00 de 1,00

Marcar questão

**Texto da questão**

Considere a área de Digital Analytics de uma empresa varejista que está iniciando sua operação no e-commerce, e deseja  melhorar  a eficácia de sua publicidade a partir de impressões de seus anúncios baseado no perfil do usuário. Para isso, está avaliando diversas informações coletadas a partir da navegação web, e deseja estimar a probabilidade de clique do usuário no anúncio apresentado.

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| Idade | Idade do usuário em anos |  |
|  |
| Renda\_região | Renda média da região na qual o usuário está localizado |  |
|  |
| Gênero | 1, se masculino e 0, se feminino |  |
|  |
| Tempo\_internet | Tempo gasto por dia na internet (em minutos) - obtido a partir dos dados da DMP (Data Management Platform) |  |
|  |
| Tempo\_site | Tempo gasto por dia no site da varejista (em minutos) - obtido com o antigo pixel instalado no site |  |
|  |
| Clicou | Flag se o usuário clicou no anúncio (1) ou não clicou (0) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de clique no anúncio - covariáveis: Idade, Renda\_região, Gênero, Tempo\_internet e Tempo\_site.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z) |
|  |
| (Intercept) | 28,51771 | 3,49556 | 8158,00000 | <0,001 |  |
| Tempo\_site | -0,20172 | 0,02647 | -7,62100 | <0,001 |  |
| Idade | 0,17601 | 0,03093 | 5,69000 | <0,001 |  |
| Randa\_região | -0,00013 | 0,00002 | -6,24500 | <0,001 |  |
| Tempo\_internet | -0,06931 | 0,00896 | -7,73900 | <0,001 |  |
| Gênero | -0,20946 | 0,50116 | -0,41800 | 0,6760 |  |

Número de observações = 700.

Avalie a tabela de estimação do modelo e escolha a alternativa correta.

Escolha uma:

a. Além do intercepto, apenas a variável explicativa Idade deve compor o modelo final, já que o coeficiente estimado é positivo.

b. Apenas a variável Gênero deve compor o modelo final, já que o p-valor = 0,676 > 5%.

c. A variável Gênero não deve compor o modelo final, já que o p-valor = 0,676 > 5%.

d. Como o coeficiente estimado para o intercepto é maior que 1, ele não deve fazer parte do modelo final.

**Feedback**

Sua resposta está correta.

A resposta correta é: A variável Gênero não deve compor o modelo final, já que o p-valor = 0,676 > 5%..

**Questão 24**

Correto

Atingiu 1,00 de 1,00

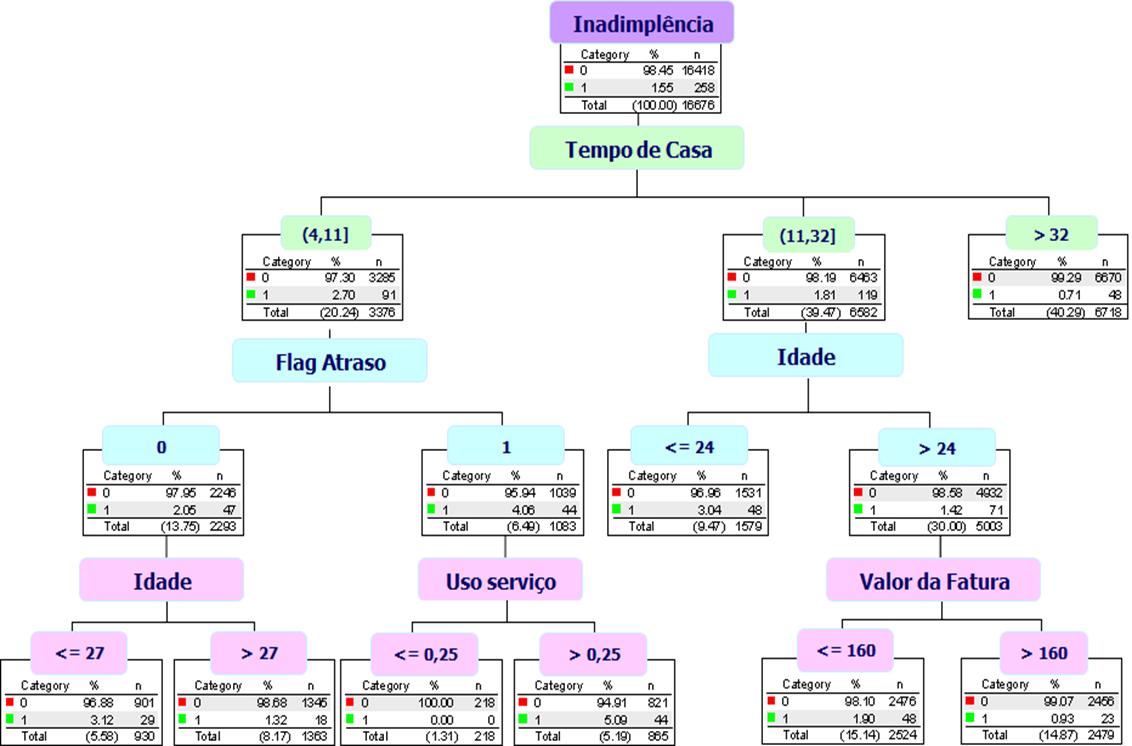
Marcar questão

**Texto da questão**

**Uma fintech com atuação exclusiva no estado do Espírito Santo deseja avaliar a inadimplência esperada de seus clientes no cenário de crise/pós-crise da pandemia de COVID-19. Para isso, reuniu as informações dos últimos 6 meses de seus clientes (Março 2020 a Agosto de 2020) e, com a ajuda de uma consultoria estatística, desenvolveu um modelo de Árvore de Decisão utilizando as seguintes informações:**

* **Inadimplência: atraso ou não pagamento de, pelo menos, uma fatura do cartão no horizonte de 3 meses (inadimplência = 1, e adimplência = 0);**
* **Tempo de Casa: tempo de relacionamento entre o cliente e a fintech (em meses);**
* **Idade: idade do cliente (em anos);**
* **Flag Atraso: flag se o cliente possui atraso no mercado, oriundo de um Bureau de Crédito;**
* **Uso serviço: percentual de vezes em que o cliente utilizou milhas para pagamento de contas no aplicativo, nos últimos 3 meses;**
* **Valor da Fatura: valor médio das últimas 3 faturas do cliente (em R$).**

**Em cada nó da Árvore de Decisão, no diagrama a seguir, o par de valores abaixo de "%" representam o percentual de clientes adimplentes (0) e inadimplentes (1) dentro do respectivo nó. Os dois valores abaixo de "n" representam as quantidades absolutas de clientes adimplentes e inadimplentes, respectivamente. Já os números da última linha, dentro de cada nó, correspondem à sua representatividade na base de dados, em percentual e em número absoluto.**



Considerando os clientes com menor tempo de casa, qual o perfil menos propenso ao evento de interesse?

Escolha uma:

a. O perfil menos propenso é formado por clientes que não possuem atraso no mercado e tem idade maior ou igual a 27 anos.

b. O perfil menos propenso é formado por clientes que possuem atraso no mercado e fizeram utilização de milhas para pagamento das contas em menos de 25% das vezes.

c. O perfil menos propenso é formado por clientes que possuem atraso no mercado e fizeram utilização de milhas para pagamento das contas em menos de 50% das vezes.

d. O perfil menos propenso é formado por clientes que não possuem atraso no mercado e não fizeram utilização de milhas para pagamento das contas em menos de 25% das vezes.

**Feedback**

Sua resposta está correta.

A resposta correta é: O perfil menos propenso é formado por clientes que possuem atraso no mercado e fizeram utilização de milhas para pagamento das contas em menos de 25% das vezes..

**Questão 25**

Incorreto

Atingiu 0,00 de 1,00

Marcar questão

**Texto da questão**

Uma empresa de concessão de crédito tem o intuito de avaliar a probabilidade de que seus clientes não cumpram com seus compromissos de pagamentos (probabilidade de default). Por meio de uma base de dados com 2.000 clientes que obtiveram crédito recentemente, a empresa pretende ajustar um modelo de regressão logística binária, utilizando como variáveis explicativas a idade, o sexo (feminino = 0; masculino = 1) e a renda mensal (R$) de cada indivíduo. A variável dependente se refere ao default propriamente dito (não default = 0; default = 1).

Output do modelo de Regressão Logística estimado para o problema de concessão de crédito - covariáveis: idade, sexo e renda.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | 2,9751 | 0,2623 | 11,34 | < 0,001 |  |
| idade | -0,0243 | 0,0070 | -3,49 | < 0,001 |  |
| sexo\_masc | 0,7415 | 0,1135 | 6,53 | < 0,001 |  |
| renda | -0,00026 | 0,0000 | -15,03 | < 0,001 |  |

Número de observações = 2000

Qual a probabilidade de default aproximada de um indivíduo do sexo feminino, com 37 anos e com renda mensal de R$ 8.650,00?

Escolha uma:

a. 70%

b. 20%

c. 85%

d. 45%

**Feedback**

Sua resposta está incorreta.

A resposta correta é: 45%.

**Questão 26**

Não respondido

Vale 1,00 ponto(s).

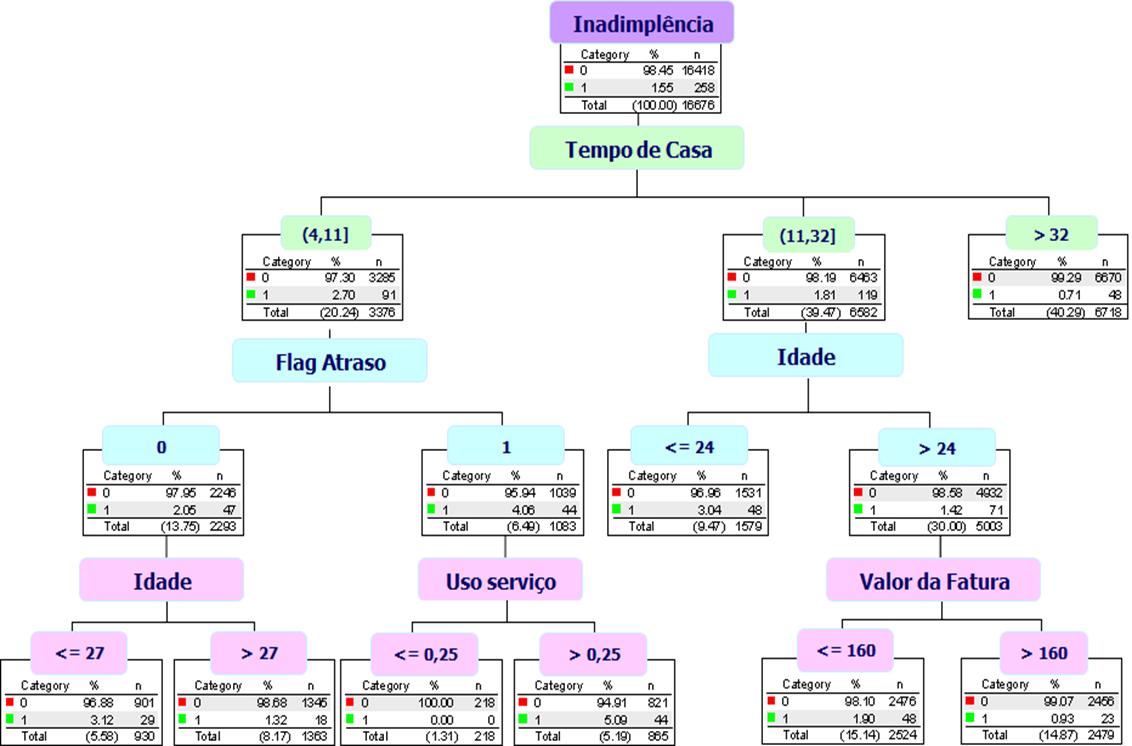
Marcar questão

**Texto da questão**

**Uma fintech com atuação exclusiva no estado do Espírito Santo deseja avaliar a inadimplência esperada de seus clientes no cenário de crise/pós-crise da pandemia de COVID-19. Para isso, reuniu as informações dos últimos 6 meses de seus clientes (Março 2020 a Agosto de 2020) e, com a ajuda de uma consultoria estatística, desenvolveu um modelo de Árvore de Decisão utilizando as seguintes informações:**

* **Inadimplência: atraso ou não pagamento de, pelo menos, uma fatura do cartão no horizonte de 3 meses (inadimplência = 1, e adimplência = 0);**
* **Tempo de Casa: tempo de relacionamento entre o cliente e a fintech (em meses);**
* **Idade: idade do cliente (em anos);**
* **Flag Atraso: flag se o cliente possui atraso no mercado, oriundo de um Bureau de Crédito;**
* **Uso serviço: percentual de vezes em que o cliente utilizou milhas para pagamento de contas no aplicativo, nos últimos 3 meses;**
* **Valor da Fatura: valor médio das últimas 3 faturas do cliente (em R$).**

**Em cada nó da Árvore de Decisão, no diagrama a seguir, o par de valores abaixo de "%" representam o percentual de clientes adimplentes (0) e inadimplentes (1) dentro do respectivo nó. Os dois valores abaixo de "n" representam as quantidades absolutas de clientes adimplentes e inadimplentes, respectivamente. Já os números da última linha, dentro de cada nó, correspondem à sua representatividade na base de dados, em percentual e em número absoluto.**



Qual o perfil menos propenso ao evento resposta? Qual a representatividade desse perfil? Qual o percentual de resposta desse perfil?

Escolha uma:

a. O perfil menos propenso é formado por clientes cujo tempo de relacionamento está entre (4, 11] meses, possuem atraso no mercado e fizeram utilização de milhas para pagamento das contas em menos de 25% das vezes. A representatividade desse perfil é 1,31%, cuja taxa de inadimplência observada foi de 100%.

b. O perfil menos propenso é formado por clientes cujo tempo de relacionamento está entre (4, 11] meses, não possuem atraso no mercado e não fizeram utilização de milhas para pagamento das contas em menos de 25% das vezes. A representatividade desse perfil é 1,31%, cuja taxa de inadimplência observada foi de 0,00%.

c. O perfil menos propenso é formado por clientes cujo tempo de relacionamento está entre (4, 11] meses, possuem atraso no mercado e fizeram utilização de milhas para pagamento das contas em menos de 25% das vezes. A representatividade desse perfil é 1,31%, cuja taxa de inadimplência observada foi de 0,00%.

d. O perfil menos propenso é formado por clientes cujo tempo de relacionamento está entre (4, 11] meses, não possuem atraso no mercado e têm idade maior que 27 anos. A representatividade desse perfil é 3,12%, cuja taxa de inadimplência observada foi de 5,58%.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: O perfil menos propenso é formado por clientes cujo tempo de relacionamento está entre (4, 11] meses, possuem atraso no mercado e fizeram utilização de milhas para pagamento das contas em menos de 25% das vezes. A representatividade desse perfil é 1,31%, cuja taxa de inadimplência observada foi de 0,00%..

**Questão 27**

Não respondido

Vale 1,00 ponto(s).

Marcar questão

**Texto da questão**

Não pode ser considerado um problema de regressão logística o seguinte caso:

Escolha uma:

a. Predizer a quantidade de chamados finalizados com sucesso no SAC de uma empresa de serviços com base na quantidade de atendentes contratados.

b. Identificar a probabilidade de uma pessoa que ainda não é cliente da instituição financeira se tornar inadimplente ao adquirir um crédito consignado.

c. Identificar os clientes com maior propensão a migração de um plano controle para um plano pós pago.

d. Identificar a probabilidade de um indivíduo sofrer um ataque cardíaco após uma cirurgia de risco.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: Predizer a quantidade de chamados finalizados com sucesso no SAC de uma empresa de serviços com base na quantidade de atendentes contratados..

**Questão 28**

Não respondido

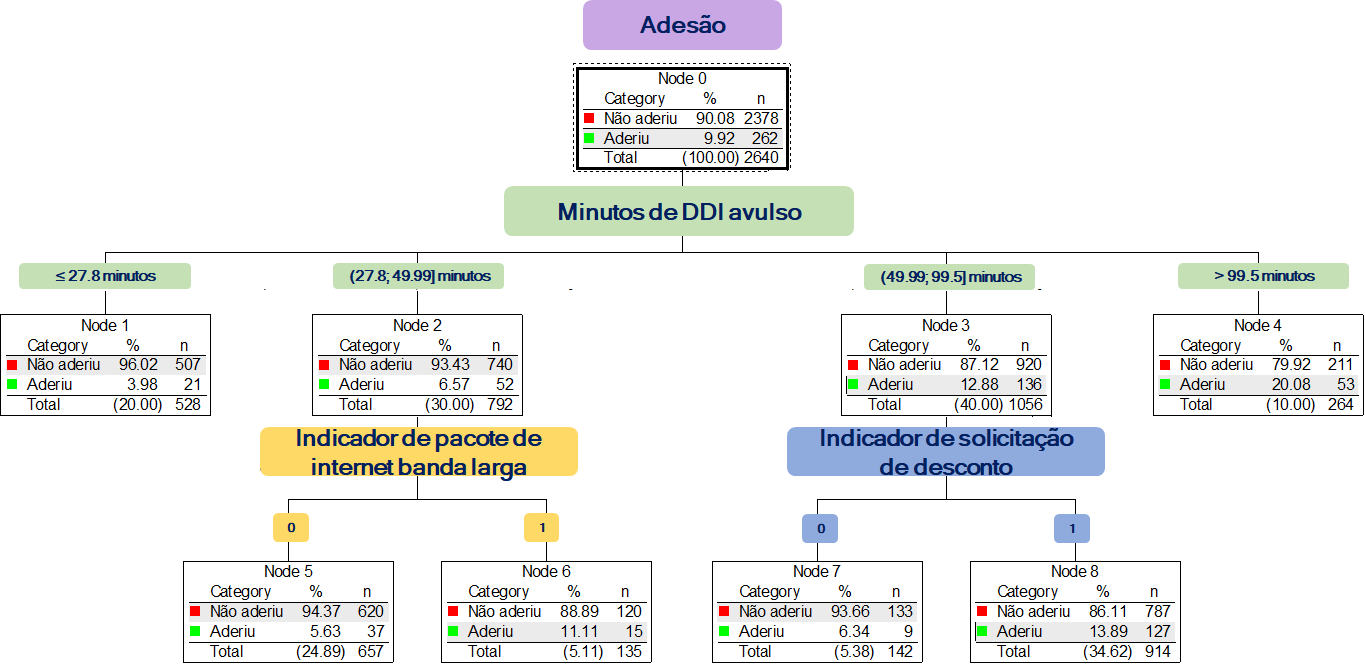
Vale 1,00 ponto(s).

Marcar questão

**Texto da questão**

Uma companhia telefônica identificou em sua base de clientes um perfil de empresas de pequeno porte que, comparado com a utilização histórica, estavam utilizando com menor frequência o serviço de chamadas internacionais. O analista de negócios responsável pela análise atribuiu essa queda de consumo aos novos softwares de conferência remota, amplamente divulgados nos últimos meses. Sendo assim, com o intuito de reter essa demanda, ofereceu a essas empresas um pacote mensal de DDI com melhores condições de contratação. A árvore abaixo representa a adesão a esse pacote no primeiro mês de oferta:

* Minutos de DDI avulso: média de minutos por dia em chamadas DDI nos últimos 3 meses.
* Indicador de pacote de internet banda larga: indicador de posse de pacote de internet banda larga na empresa.
* Indicador de solicitação de desconto: indicador de solicitação de desconto pelo cliente nos últimos 3 meses em algum dos serviços contratados pelo cliente.



Do total de nós finais, quantos nós são classificados como propensos ao evento de interesse (resposta 1)?

Escolha uma:

a. Dada a proporção de adesão da base (90,08%), nenhum perfil é identificado com propensão ao evento de interesse.

b. Dada a proporção de adesão geral da base (9,92%), identificamos 6 perfis com propensão ao evento de interesse.

c. Dada a proporção de adesão geral da base (9,92%), identificamos 2 perfis com propensão ao evento de interesse.

d. Dada a proporção de adesão geral da base (9,92%), identificamos 3 perfis com propensão ao evento de interesse.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: Dada a proporção de adesão geral da base (9,92%), identificamos 3 perfis com propensão ao evento de interesse..

**Questão 29**

Não respondido

Vale 1,00 ponto(s).

Marcar questão

**Texto da questão**

Qual das seguintes medidas estatísticas apresentadas resume, de forma razoável, o nível de acerto de um modelo de regressão logística?

Escolha uma:

a. A sensibilidade, sendo que quanto menor o valor dessa medida, maior o nível de discriminação do modelo entre as classes de respostas.

b. O KS (Kolmogorov-Smirnov), sendo que quanto menor o valor dessa medida, maior o nível de discriminação do modelo entre as classes de respostas.

c. O KS (Kolmogorov-Smirnov), sendo que quanto maior o valor dessa medida, maior o nível de discriminação do modelo entre as classes de respostas.

d. A especificidade, sendo que quanto maior o valor dessa medida, maior o nível de discriminação do modelo entre as classes de respostas.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: O KS (Kolmogorov-Smirnov), sendo que quanto maior o valor dessa medida, maior o nível de discriminação do modelo entre as classes de respostas..

**Questão 30**

Não respondido

Vale 1,00 ponto(s).

Marcar questão

**Texto da questão**

Considere que o departamento de RH,  por meio da sua área de People Analytics, tenha o interesse em investigar as variáveis explicativas  que influenciam na probabilidade de um trainee chegar atrasado à empresa. Sendo assim, o RH elaborou uma pesquisa com os 100 trainees perguntando se ele costuma ou não chegar atrasado no trabalho. Perguntou também sobre a distância percorrida no trajeto (em quilômetros), o número de semáforos pelos quais cada um passou, o período em que foi realizado o trajeto (manhã ou tarde) e como cada um se considera em termos de perfil ao volante (calmo, moderado ou agressivo).

|  |  |
| --- | --- |
| Variável | Descrição |
|  |
| tempo | Tempo para chegar à empresa (em minutos) |  |
|  |
| dist | Distância percorrida até a empresa (quilômetros) |  |
|  |
| sem | Quantidade de semáforos |  |
|  |
| per | Período do dia no qual o trainee foi à empresa. Manhã (categoria de referência - dummy = 1): caso o trainee tenha ido mais cedo para atividades específicas de sua área ou Tarde (dummy = 0): se o trainee foi apenas para o período da tarde, horário definido para as atividades pertinentes ao programa de Trainee |  |
|  |
|  |
| perfil | Perfil ao volante (Perfil 1: calmo, Perfil 2: moderado e Perfil 3: agressivo) |  |
|  |

Output do modelo completo de Regressão Logística estimado para o problema de atraso do trainee - covariáveis: dist, sem, per e perfil.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | -30,2003 | 9,9811 | -3,03 | 0,002 |  |
| dist | 0,2202 | 0,1097 | 2,01 | 0,045 |  |
| sem | 2,7667 | 0,9217 | 3,00 | 0,003 |  |
| per | -3,6534 | 0,8781 | -4,16 | <0,001 |  |
| perfil\_2 | 1,3460 | 0,7477 | 1,80 | 0,072 |  |
| perfil\_3 | 2,91447 | 1,1788 | 2,47 | 0,013 |  |

Número de observações = 100

Output obtido via stepwise do modelo de Regressão Logística estimado para o problema de atraso do trainee - covariáveis: dist, sem, per e perfil.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Coeficiente | Erro padrão | Valor z | Pr(>|z|) |
|  |
| Intercepto | -30,9334 | 10,6363 | -2,91 | 0,004 |  |
| dist | 0,2041 | 0,1012 | 2,02 | 0,044 |  |
| sem | 2,9201 | 1,0108 | 2,89 | 0,004 |  |
| per | -3,7763 | 0,8467 | -4,46 | <0,001 |  |
| perfil\_3 | 2,45907 | 1,1395 | 2,16 | 0,031 |  |

Considere a análise de desempenho apresentada a seguir e indique a alternativa correta com relação ao desempenho do modelo estimado:

|  |  |  |  |
| --- | --- | --- | --- |
| Observado | Estimado | | Total |
| Não atraso | Atraso |
| Não atraso | 28 | 13 | 41 |
| Atraso | 2 | 57 | 59 |
| Total | 30 | 70 | 100 |

|  |  |
| --- | --- |
| Sensibilidade | 96,61% |
| Especificidade | 68,29% |

Escolha uma:

a. A proporção de casos de não atraso corretamente classificados pelo modelo é 96,61% e o acerto global do modelo é 85,00%.

b. A proporção de casos de atraso corretamente classificados pelo modelo é 96,61% e o acerto global do modelo é 85,00%.

c. A proporção de casos de não atraso corretamente classificados pelo modelo é 68,29% e o acerto global do modelo é 96,61%.

d. A proporção de casos de atraso corretamente classificados pelo modelo é 68,29% e o acerto global do modelo é 96,61%.

**Feedback**

Sua resposta está incorreta.

A resposta correta é: A proporção de casos de não atraso corretamente classificados pelo modelo é 96,61% e o acerto global do modelo é 85,00%..

Parte inferior do formulário