

Sobre o problema: um famoso empresário de jovens talentos da NBA está buscando otimizar sua busca por jovens estrelas na NBA. Para isso, ele o contrata para criar um algoritmo que classifica os atletas como promissores a uma grande carreira ou não, baseado em suas estatísticas no seu primeiro ano de NBA. Ele considera que uma jovem promessa atingiu uma grande carreira quando ele consegue permanecer 5 anos ou mais na liga NBA.

Sobre o banco de dados: o banco de dados possui as estatísticas de 1340 jovens, obtidas durante seu ano de estreia na NBA. São 21 variáveis, das quais 1 é para identificação (Name), 19 são explicativas e 1 é a variável resposta (TARGET_5Yrs), que é 1 se o jogador ficou 5 anos ou mais na liga NBA, ou 0 caso contrário. Em anexo, segue tabela com os detalhes de cada variável. Para acessar o banco de dados:

```
dados <- read.csv("https://query.data.world/s/6uhvcowok54gcj5jz3i2ccwldlb6rl",  
header=TRUE, stringsAsFactors=FALSE)
```

Sobre a divisão treino/teste: o cliente exige que 25% das amostras sejam usadas como amostra teste, para uma boa estimativa da acurácia, sensibilidade e especificidade. Utilize seu número de matrícula como set.seed().

Sobre tratamento dos dados: qualquer método de pré-processamento pode ser utilizado, mas toda metodologia utilizada deve ser deixada bem claro no relatório final.

Sobre o treinamento : deve-se utilizar o Gradiente Boosting. Para definir os valores dos hiperparâmetros, devem ser testados PELO MENOS 5 valores diferentes para profundidade da árvore, 3 para a taxa de aprendizado e 4 para o número de árvores. Alterações no n.minobsinnode são opcionais. A utilização das funções train() ou gbm() ficam à sua escolha. O critério de seleção final do modelo deve ser o com maior AUC. Utilize seu número de matrícula como set.seed().

Sobre o relatório final: deve ser um relatório final, em formato de apresentação (beamer), a ser apresentado ao contratante. Esse relatório deve conter (de maneira explicativa) toda metodologia empregada, resultados dos modelos, modelo final e comportamento do modelo final na amostra teste.

O que deve ser entregue na data final: o relatório (.pdf) e os códigos utilizados (.txt). Entre os dias 15 e 17 de setembro teremos as apresentações (horários a definir em breve).

	Description
Name	Name
GP	Games Played
MIN	MinutesPlayed
PTS	PointsPerGame
FGM	FieldGoalsMade
FGA	FieldGoalAttempts
FG%	FieldGoalPercent
3P Made	3PointMade
3PA	3PointAttempts
3P%	3PointAttempts
FTM	FreeThrowMade
FTA	FreeThrowAttempts
FT%	FreeThrowPercent
OREB	OffensiveRebounds
DREB	DefensiveRebounds
REB	Rebounds
AST	Assists
STL	Steals
BLK	Blocks
TOV	Turnovers
TARGET_5Yrs	Outcome: 1 if career length >= 5 yrs, 0 if < 5...