

Estatística Preditiva: Modelo Logístico para predição de rebaixamento de times na Série A do campeonato brasileiro de futebol.

Josué Filipe Pontes Nery; Marcela Cristyne; Thaís Cordeiro Pereira

Objetivos

A proposta, aqui apresentada, é de elaborar um modelo matemático que possa prever a probabilidade de um clube ser rebaixado ou não. Para tal fim, foi analisada a performance dos clubes que participaram do campeonato brasileiro de futebol (Série A), de 2010 a 2020, considerando as seguintes variáveis: gols sofridos, gols marcados e pontuação final.

Metodologia

O método de Regressão Logística foi escolhido pois possui a capacidade de elaborar um modelo matemático que mostra a relação das variáveis auxiliares, sua influência na variável de resposta (rebaixado) e a predição do rebaixamento de um clube. Para montar a base de dados foram selecionadas as informações dos clubes rebaixados (quatro clubes) e dos quatro clubes que “fugiram” do rebaixamento, de cada ano.

Resultados

Os resultados obtidos tiveram certa consistência, considerando uma base de dados com oitenta e oito (88) elementos. Temos uma taxa de acerto média de 88,63636%, e ao observarmos a matriz de confusão do modelo temos uma taxa de acerto para clubes não rebaixados de 95,4545% e de 81,8181% para times rebaixados. Já ao analisarmos a curva ROC (receiver operating characteristic), obtivemos a área abaixo da curva ROC, ou “AUC-value”, com 0.9607438.

Por fim, observamos uma forte correlação entre as variáveis de gols marcados e pontuação final, gols sofridos e pontuação final, porém uma fraca relação entre gols marcados e gols sofridos. O modelo possui um certo nível de confiabilidade considerando a base de dados trabalhada.