



Métodos Estatísticos Aplicados na Modelagem de Odds Ratio em Apostas Desportivas.

Caio Henrique R. G. de Medeiros¹,Alexsandro Bezerra Cavalcanti ²

RESUMO

As apostas esportivas tornaram-se cada vez mais comuns, principalmente com o advento de casas de apostas online. Desta forma, o uso conjunto de modelos estatísticos e algoritmos computacionais para analisar e tentar prever resultados nestes cenários também sofreu grande aumento. Este trabalho teve como objetivo específico realizar previsões de vitória ou derrota/empate do mandante em apostas esportivas de Futebol, através da aplicação de 03 algoritmos de aprendizado de máquina diferentes – Naive Bayes, K-Vizinhos e Regressão Logística - e comparar suas acurácias na previsão dos resultados. O projeto foi realizado a partir do entendimento dos modelos estatísticos e dos algoritmos selecionados e de sua adequação e aplicação ao foco do projeto, com todos os códigos sendo construídos na linguagem de programação Python3. A base de dados utilizada foi composta por tabelas do campeonato de futebol inglês, a Premier League, dos anos de 2012 a 2018. Os algoritmos produziram uma Acurácia média, Sensibilidade, oddH Média e Lucro ou Prejuízo total. Os resultados foram positivos, pois os 3 programas apresentaram lucro positivo ao final dos testes, tendo os modelos de Naive Bayes e Regressão Logística apresentado maiores Acurácia e Sensibilidade. A partir dos resultados, pôde-se concluir que o algoritmo de Regressão Logística, apesar de não ter retornado o maior Lucro total, demonstrou-se ser mais preciso e seguro em suas previsões.

Palavras-chave: odd, futebol, desportiva, aposta.

¹Aluno de Ciência da Computação, DSC, UFCEG, Campina Grande, PB, e-mail: caio.medeiros@ufcg.edu.br

²Doutor, Professor, UAEst, UFCEG, Campina Grande, PB, e-mail: alexsandro.bezerra@ufcg.edu.br



Statistical Methods Applied to Odds Ratio Modeling in Sports Betting.

ABSTRACT

Sports betting has become increasingly common, especially with the creation of online bookmakers. Thus, the joint use of statistical models and computational algorithms to analyze and try to predict results in these scenarios has also increased greatly. This work had the specific objective of making predictions of victory or defeat / tie of the client in soccer sports betting, through the application of 03 different machine learning algorithms - Naive Bayes, K-Next Neighbors and Logistic Regression - and to compare their accuracy in the forecast results. The project was carried out from the understanding of the selected statistical models and algorithms and their suitability and application to the project's focus, with all the codes being built in the Python3 programming language. The database used was composed of excel datasheets from the English soccer championship, the Premier League, from 2012 to 2018. The algorithms produced an average Accuracy, Sensitivity, Average oddH and Total Profit or Loss for each year. The results were positive, as the 3 programs showed a positive profit at the end of the tests, with the Naive Bayes and Logistic Regression models showing greater Accuracy and Sensitivity. From the results, it was possible to conclude that the Logistic Regression algorithm, in spite of not having returned the highest total Profit, proved to be more accurate and safe in its forecasts.

Keywords: odd, soccer, sports, betting.