



USANDO TÉCNICAS DE MINERAÇÃO DE DADOS PARA PREVER RESULTADOS DE PARTIDAS DE FUTEBOL

Lucas Barros Rocha¹, Carlos Eduardo Santos Pires²

RESUMO

Com a quantidade de dados disponibilizados na Internet nos últimos anos relacionados ao futebol, surgiu o interesse de conhecer mais sobre o esporte e tentar prever resultados de partidas a partir do conhecimento adquirido com resultados históricos. Este trabalho propõe um método para predição de resultados de partidas de futebol utilizando mineração de dados. Para tal, foram feitas extrações de dados a partir de portais online e sites de casas de apostas que oferecem dados publicamente, para formação de uma base de dados integrada. Em seguida, foi aplicado um processo de limpeza, organização e derivação de dados, para que algoritmos de aprendizagem de máquina pudessem buscar padrões implícitos nos dados, com o intuito de obter o melhor valor de predição. Foram alcançados acertos superiores aos das casas de apostas as quais, por si só, podem ser consideradas boas preditoras. Na predição de vitória do time da casa, por exemplo, foi obtido um valor de predição 10,71% superior ao das casas de apostas.

Palavras-chave: futebol, apostas, algoritmos, predição, casas de apostas.

¹ Aluno do Curso de Ciência da Computação, Departamento de Sistemas e Computação, UFCG, Campina Grande, PB, e-mail: lucas.rocha@ccc.ufcg.edu.br

² Doutor em Computação, Professor Associado I, Departamento de Sistemas e Computação, UFCG, Campina Grande, PB, e-mail: cesp@computacao.ufcg.edu.br

USING DATA MINING TECHNIQUES TO PREDICT FOOTBALL MATCH RESULTS

ABSTRACT

With the amount of data available on the Internet in recent years related to football, the interest to know more about the sport and try to predict results of matches based on the knowledge acquired with historical results has increased. This work proposes a method for predicting soccer match results using data mining. To this end, data extractions were made from online portals and bookmaker sites that offer data publicly in order to construct an integrated database. Then, a process of data cleansing, organization and derivation was applied, in such a manner that machine learning algorithms could search for patterns implicit in the data, in order to obtain the best prediction value. We have achieved higher scores than the ones provided by bookmakers which, by themselves, can be considered good predictors. For instance, in the prediction of victory of the home team, we achieved a prediction value obtained 10.71% higher than that the one of bookmakers.

Keywords: football, betting, algorithms, prediction, bookmakers.