O banco de dados final do futebol para análise de dados e aprendizagem de máquinas

O que você obtém:

* +25,000 jogos
* +10.000 jogadores
* 11 países europeus com seu campeonato de liderança
* Épocas 2008 a 2016
* Atributos de jogadores e equipes \* provenientes da série de videogames FIFA da EA Sports, incluindo as atualizações semanais
* Equipe se alinha com formação de esquadrão (coordenadas X, Y)
* Probabilidades de apostas de até 10 provedores
* Eventos de correspondência detalhados (tipos de gols, posse, canto, cruz, faltas, cartas, etc ...) para mais de 10.000 jogos

\* *16 de outubro de 2016: Nova tabela contendo os atributos das equipes da FIFA!*

Fonte de dados original:

Você pode facilmente encontrar dados sobre partidas de futebol, mas eles geralmente estão espalhados por diferentes sites. Uma completa coleta e processamento de dados foi feito para tornar sua vida mais fácil. Devo insistir que você não faça nenhum uso comercial dos dados . Os dados foram obtidos de:

* <http://football-data.mx-api.enetscores.com/> : pontuação, formação, formação de equipes e eventos
* <http://www.football-data.co.uk/> : apostas. [Clique aqui para entender o sistema de nomeação de colunas para apostas:](http://www.football-data.co.uk/)
* <http://sofifa.com/> : atributos de jogadores e equipes dos jogos EA Sports FIFA. [Série FIFA e todos os ativos da FIFA ativos da EA Sports.](http://sofifa.com/)

Quando você verifica o banco de dados, você notará chaves estrangeiras para jogadores e as correspondências são as mesmas que as origens de dados originais. Liguei para essas chaves estrangeiras "api\_id".

Melhorando o conjunto de dados:

Você notará que alguns jogadores estão ausentes da programação (valores NULL). Isso ocorre porque eu não consegui obter seus atributos da FIFA. Isso será corrigido horas extras à medida que o algoritmo de rastreamento está sendo melhorado. O conjunto de dados também será expandido para incluir jogos internacionais, copas nacionais, Liga dos Campeões e Liga Europa. Pergunte-me se você está após um torneio específico.

Entre em contato comigo se quiser ajudar a melhorar este conjunto de dados.

[CLIQUE AQUI PARA ACESSAR O PROJETO GITHUB](https://github.com/hugomathien/football-data-collection/tree/master/footballData)

*Nota importante para as pessoas interessadas em usar os rastreadores:* desde que eu escrevi pela primeira vez os scripts de rastreamento (em python), parece que o sofifa.com mudou seu design e com isso vem novos requisitos para os scripts. *O script existente para rastrear os jogadores ('Player Spider') não funcionará até eu atualizá-lo.*

Explorando os dados:

Agora, essa é a parte divertida, há muito que você pode fazer com este conjunto de dados. Eu estarei adicionando visuais e insights para esta página de visão geral, mas por favor dê uma olhada nos kernels e experimentá-lo! Aqui estão algumas idéias para você:

O Santo Graal ... ... é obviamente para prever o resultado do jogo. Os bookies usam 3 classes (Home Win, Draw, Away Win). Eles conseguem cerca de 53% das vezes. Isso também é o que eu consegui até agora usando meu próprio SVM. Embora pareça alto para um jogo de esporte tão aleatório, você deve saber que o time da casa ganha cerca de 46% do tempo. Portanto, o caso base (Previsão constante de Home Win) tem precisão de 46%.

Probabilidades vs Odds

Ao executar um classificador multi-classe como SVM, você também pode produzir uma estimativa de probabilidade e compará-lo com as chances de apostas. Dê uma olhada em suas probabilidades de contrapartida e veja quais jogos você teve previsões muito diferentes.

Explore e visualize os recursos

Com acesso aos atributos de jogadores e equipes, formações de equipe e eventos no jogo você deve ser capaz de produzir algumas idéias interessantes sobre [The Beautiful Game](https://en.wikipedia.org/wiki/The_Beautiful_Game) . Quem sabe, o próprio Guardiola pode contratar um de vocês algum dia!

URL page: https://www.kaggle.com/hugomathien/soccer