

UNIVERSIDADE POSITIVO

CRISTIANO FONTANA

JAN ANTONIO PEREIRA

LUCAS FLECK DE SOUZA CASTRO

Predição no Futebol usando Machine Learning

CURITIBA

2017

Tema

Diante de um cenário de alto interesse do mercado corporativo e a forte carência de profissionais nas áreas de Machine Learning e Data Science, devido a enorme curva de aprendizado, resolvemos então utilizar um assunto de domínio e conhecimento popular como o futebol para tornar mais leve o aprendizado e a implementação de técnicas de análise de dados, aprendizado de máquina e processamento distribuído com o objetivo de prever eventos dentro de uma partida de futebol utilizando como modalidade de projeto uma monografia.

Problema

A previsão de eventos dentro de uma partida de futebol é algo extremamente difícil, devido à grande quantidade de variáveis que podem ou não estar relacionadas.

Dentro de uma partida de futebol tudo o que ocorre em uma partida representa um evento e no mundo dos esportes todo o evento dentro de uma partida pode se tornar um mercado em potencial, por exemplo placar exato, número de gols, número de cartões, número de escanteios, etc.

Todo o mercado e respectivo evento é medido pelo que chamamos de ODDs, nada mais é que um fator matemático que traduz a probabilidade do evento ocorrer.

Diferentes métodos foram propostos ao longo dos anos sobre como prever uma partida de futebol. Alguns métodos utilizam apenas dados quantitativos outros apenas dados qualitativos, enquanto outros usam uma mistura de ambos quantitativos e qualitativos na previsão de partidas de futebol. Um conjunto de dados quantitativos seria um conjunto de dados que consistiria em dados históricos como resultados diretos, performances ou chutes recentes a gol, enquanto um conjunto de dados qualitativos sobre conhecimento específico a respeito do desempenho do time, jogador, disponibilidade do jogador, condições do gramado, altitude do local, opinião dos torcedores nas redes sociais ou a pressão da mídia.

Existem diversos trabalhos e estudos em cima do uso de técnicas de análise de dados no futebol, onde grande parte deles voltados a dados quantitativos com uma abrangência muito restrita.

Embora, esses métodos propõem um caminho, a utilização e comparação de vários métodos de Machine Learning para prever o resultado de uma partida de futebol não é uma abordagem tão comum para trabalhos de pesquisa de previsão de futebol onde geralmente emprega-se apenas análises estatísticas.

Objetivos

Para atingirmos nosso objetivo construiremos um conjunto de crawlers para capturar o máximo de dados quantitativos e qualitativos possíveis que considerarmos relevantes para construção de nossa base de dados.

Com a base dados carregada, faremos a limpeza, tratamento e seleção dos dados onde iremos testar um conjunto de algoritmos de machine learning e comparar seus resultados em busca de um modelo ideal de predição de um dado evento em uma partida de futebol.

Como o volume de dados e do tempo de resposta dos algoritmos tenderão a crescer com o tempo planejamos a construção da solução sob ferramentas de Big Data de forma escalável permitindo o processamento distribuído via cluster.

Resultados Esperados

Estamos bem conscientes que os resultados dos modelos que iremos construir dependerá do volume de informações relevantes que conseguiremos coletar, da escolha das variáveis e dos algoritmos de aprendizagem que farão parte da construção de nossos modelos preditivos.

Também sabemos que cada tipo de algoritmo de aprendizagem de máquina apresentará uma taxa diferente, deveremos então avaliá-los de acordo com um conjunto de indicadores como acuracidade, sensibilidade, precisão, F1 Score e Curva de ROC, entre outros indicadores.

Atualmente os artigos publicados sobre o tema tem apresentado resultados de acordo com o tipo de algoritmo de aprendizagem e uma análise específica de mercado, como por exemplo a análise restrita de campeonato mais equilibrado como a Premier League no futebol Inglês.

Com os pés no chão e muito trabalho, faremos dos nossos modelos o mais preditivo possível, mas sem dúvida o maior resultado será a experiência adquirida em análise e tratamento de dados, aplicação de técnicas de machine learning e ferramentas de Big Data.

No futuro os dados coletados poderão ser tornar APIs comerciais, incluindo serviços de predição de eventos em partidas de futebol e automatização de bots para realização de ações de Trader Esportivo entre outros.

Orientador

Em conversa com a professora de Spark e Machine Learning, Mariane Gonçalves Kulik a mesma confirmou interesse de nos orientar no decorrer deste trabalho.

Curitiba, 23 de outubro de 2017.

CRISTIANO FONTANA

JAN ANTONIO PEREIRA

LUCAS FLECK DE SOUZA CASTRO