



Data Science Academy

www.datascienceacademy.com.br

Formação Cientista de Dados

Projeto com Feedback 9

Prevendo Tendências Macroeconômicas

Como podemos usar as ferramentas e a inteligência do mundo para prever resultados econômicos que nunca podem ser totalmente previsíveis? Essa questão está no centro de inúmeras atividades econômicas em todo o mundo - inclusive na Two Sigma Investments, que vem aplicando tecnologia e estratégias sistemáticas ao comércio financeiro desde 2001.

Por mais de 15 anos, a Two Sigma tem estado na vanguarda da aplicação de tecnologia e ciência de dados às previsões financeiras. Embora seus avanços pioneiros em big data, inteligência artificial e aprendizado de máquina no mundo financeiro tenham impulsionado o mercado, como acontece com todos os outros avanços científicos, eles são levados a progredir continuamente.

A oportunidade econômica depende da capacidade de fornecer previsões singularmente precisas em um mundo de incertezas. Ao prever com precisão os movimentos financeiros, você aprenderá sobre abordagens orientadas cientificamente para desbloquear recursos preditivos significativos. A Two Sigma está animada para encontrar valor preditivo e obter uma melhor compreensão das habilidades oferecidas pela multidão global de ciência de dados.

Para este projeto, recomendamos a utilização da linguagem R, com o dataset disponível aqui:

<https://www.kaggle.com/c/two-sigma-financial-modeling/>

Este conjunto de dados contém recursos anônimos relacionados a um valor variável no tempo para um instrumento financeiro. Cada instrumento tem um id. O tempo é representado pelo recurso 'timestamp' e a variável a prever é 'y'. Nenhuma informação adicional será fornecida sobre o significado dos recursos, as transformações que foram aplicadas a eles, a escala de tempo ou o tipo de instrumentos que estão incluídos nos dados. Além disso, de acordo com as regras, os alunos não devem usar dados além dos dados vinculados ao site.

Os dados são salvos e acessados como um arquivo .h5. Usamos o formato de arquivo .h5 em vez do formato padrão .csv para obter velocidades de leitura mais rápidas. O arquivo do conjunto de treinamento está disponível para download e modelagem offline. O conjunto de testes não está disponível para download.

Quando concluir o projeto, envie os scripts e datasets para projeto@dsacademy.com.br. Caso os datasets usados sejam muito grandes, armazene em um diretório virtual (existem vários na internet, como Google Drive ou Dropbox) e envie o link para que nossa equipe possa baixar os datasets. Se os arquivos foram pequenos (uma amostra do dataset original), envie no anexo junto com o script. Documente seu script tanto quanto possível.

Caso prefira, disponibilize seu projeto no Github e envie o link do seu repositório para nossa equipe no e-mail projeto@dsacademy.com.br. Nesse caso, o Readme do repositório deve constar que este trata-se de um projeto da Formação Cientista de Dados da Data Science Academy.

Em até 24 horas, daremos o feedback!

Bom trabalho!