

Desafio DSC 2022

Fase 1

A primeira etapa basicamente compreende a predição de $k = 20$ pregões futuros dos ativos que compõem o índice SP500. O SP500 é um índice formado por 500 grandes empresas listadas em bolsas americanas composto por, curiosamente, 503 ativos. A listagem dos ativos pode ser encontrada, por exemplo, no [Wikipedia: SP500](#).

Para um dado ativo i , considere seu log-retorno de k posições atrás correspondente ao dia de pregão t como

$$r_{i,t} = \ln \frac{p_t}{p_{t-k}},$$

onde p_j é seu preço de fechamento no dia de pregão j .

Cada time terá até 21/10 para submeter o log-retorno dos dias **24/10 à 18/11**, representado aqui por $t = [1, 20]$. Ou seja, deverão ser reportados os log-retornos dos dias **[24/10, 18/11]** relativos, respectivamente, aos dias anteriores **[26/09, 21/10]**.

Não há restrições para a solução, apenas é exigido que as bases de dados utilizadas sejam públicas. Veja as fontes sugeridas ao final deste documento. Adicionalmente, as 5 melhores equipes deverão também submeter o modelo e dados de treinamento utilizados nesta primeira fase.

O ranqueamento dos times será feito com base nos erros

$$e_{i,t} = r_{i,t} - \hat{r}_{i,t}$$

entre o retorno previsto $\hat{r}_{i,t}$ e realizado $r_{i,t}$ de um ativo i no pregão t ; a variância da amostra

$$s_i^2 = \frac{1}{k-1} \sum_{t=1}^k (e_{i,t} - \bar{e}_i)^2$$

dos erros, onde \bar{e}_i é a média destes erros; e o erro quadrático médio

$$EQM_i = \frac{1}{k} \sum_{t=1}^k (e_{i,t})^2$$

entre o retorno previsto $\hat{r}_{i,t}$ e realizado $r_{i,t}$ de cada ativo i .

Os times serão ranqueados em ordem crescente de acordo com

$$E = \sum_{i=1}^{503} \sqrt{EQM_i + s_i^2}.$$

A lista final de colocados serão divulgados no dia 18/11 após o fechamento do pregão. As 5 melhores equipes seguem para Etapa 2. Divulgaremos semanalmente no site, para os primeiros k dias realizados, os resultados parciais dos times que enviarem as previsões.

Formato

Cada time deverá submeter um arquivo .csv de acordo com o modelo [‘predicao.csv’](#) disponível na pasta compartilhada [‘Submissão’](#).

O formato exige:

- colunas separadas por ‘,’ (vírgulas)
- valores com decimais utilizando ‘.’ (ponto)
- Primeira coluna com nome ‘Dia’
- Demais colunas são os tickers dos ativos da SP500 em ordem crescente
- Cada linha deve conter o **retorno** referente a um dia previsto

Para garantir que o formato está correto, foi criado um [script simples em python](#) para testar seu arquivo antes da submissão.

```
# teste antes de submeter  
python check_submission.py arquivo.csv
```

As submissões serão realizadas via Google Forms. Será fornecido um link via email, para as equipes submeterem suas previsões. Apenas a última submissão será considerada.

Importante: as submissões devem ser feitas com o mesmo email utilizado na inscrição, ou não serão válidas.

Fontes

Abaixo seguem sugestões de fonte de dados e outras informação que podem ser úteis.

- [Yahoo Finance](#)
- [Dados fundamentalistas](#)
- [Data Nasdaq](#)
- [Captura de dados em Python](#)