

Plan Overview

A Data Management Plan created using DMPTool

Title: Trabalho de conclusão de curso

Creator: Livia Almeida

Affiliation: Universidade de São Paulo (www5.usp.br)

Funder: Digital Curation Centre (dcc.ac.uk)

Template: Digital Curation Centre

Project abstract:

O projeto tem como objetivo realizar a otimização de carteiras de investimento, fazendo com que elas apresentem um retorno que segue o índice Ibovespa. Para isso, são selecionados os cinco fundos com maior representatividade dentro do índice para compor a carteira e, com a utilização de métodos de programação matemática, essa carteira deve seguir o benchmark.

Last modified: 08-30-2022

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Trabalho de conclusão de curso

Os dados que serão coletados são os valores de abertura, fechamento, máximo, mínimo, volume e variação de cada um dos ativos, incluindo o Ibovespa. Esses dados serão coletados a cada mês, para garantir uma granularidade, mas também, para que seja possível realizar a otimização.

O dado será coletado através do site investing.com. Nele, é possível coletar os dados com a granularidade desejada e de todos os anos que precisarão ser utilizados no projeto. Após a coleta, é necessário realizar o tratamento dos dados, para garantir que as colunas estejam de acordo com um padrão e os valores estejam todos formatados da mesma maneira.

As informações necessárias para que o dado seja lido e interpretado são:

- Qual é o benchmark
- Quais são os fundos que irão compor a carteira
- Qual período será utilizado para realizar a otimização
- Qual a granularidade dos dados

Todas essas informações são definidas no momento de coleta e tratamento dos dados e são documentadas ao longo da descrição de cada simulação.

Os dados são públicos e, portanto, não terão problemas em relação à sua utilização. Além disso, os dados apenas serão manuseados pelos integrantes do grupo e, com o compartilhamento das análises, sempre será priorizado o cuidado com as fontes utilizadas e a veracidade de todas as informações.

Como os dados são públicos, não será necessário lidar com essas questões.

Durante as pesquisas, os dados que estiverem sendo utilizados serão mantidos no repositório do projeto. Como não será uma quantidade tão grande de informações, não será necessário nenhum tipo de armazenamento ou cobranças adicionais.

Para garantir a segurança de acesso, apenas os membros do grupo poderão editar e manusear os dados coletados e tratados.

Todos os dados utilizados são atualizados de acordo com o período. A cada dia novos valores são atualizados na bolsa e, portanto, é necessário coletar os dados novamente, caso se tenha o objetivo de fazer uma otimização mais condizente com a realidade atual.

Entretanto, o período que será utilizado é uma decisão de projeto, que pode ser mantido enquanto durarem as análises.

O plano para a preservação dos dados é que eles sempre se mantenham atualizados no repositório do projeto. A cada novo período definido, um novo conjunto de dados deve ser coletado e salvo em uma pasta que irá conter todos os dados utilizados no projeto.

Como não é uma quantidade tão grande de dados, a sua preservação no repositório se torna possível e livre de custos.

Não será necessário compartilhar o dado, apenas caso a pessoa tenha interesse em observar o repositório do projeto. O dado será apenas utilizado entre os integrantes do grupo que serão responsáveis por fazer as otimizações e simulações. Além dessas pessoas, a única coisa que efetivamente precisará ser compartilhada serão as análises finais.

Como o dado não precisará ser compartilhado, não há nenhuma restrição.

Os responsáveis pelo gerenciamento dos dados serão os três integrantes do grupo do trabalho de conclusão de curso.

Para entregar o plano, será necessário o acesso ao repositório do GitHub que irá conter os dados (https://github.com/diego-arruda/TCC_DATA), conhecimento de Python para a realização do tratamento e criação das tabelas, e conhecimento de Matlab para a realização dos códigos que irão implementar as simulações.
