

Bootcamp: Analista de Dados com ênfase para o Mercado Financeiro

Desafio Prático

Módulo 2: Fundamentos de Programação para Ciência de Dados

Objetivos de Ensino

Exercitar os seguintes conceitos trabalhados no Módulo:

1. Familiarizar-se com a importação e manipulação de conjuntos de dados do mercado financeiro usando Python.
2. Analisar e interpretar tendências e padrões em dados históricos de ações ou outros instrumentos financeiros.
3. Visualizar as tendências de mercado, variações de preços e volume de negociações ao longo do tempo.
4. Implementar e testar modelos de aprendizado de máquina para previsão de movimentos futuros no mercado.
5. Desenvolver habilidades práticas em bibliotecas Python essenciais para ciência de dados, focando em aplicações no mercado financeiro.

Enunciado

O mercado de capitais é um ambiente altamente dinâmico, no qual os preços dos ativos financeiros podem variar significativamente em curtos períodos. Como cientista de dados com ênfase no mercado financeiro, você foi desafiado a analisar um conjunto de dados contendo informações sobre ações de uma determinada empresa listada na bolsa de valores. Seu objetivo é não apenas entender as tendências passadas, mas também criar um modelo preditivo para auxiliar nas decisões de investimento. Leia as atividades, premissas, nota e sugestão dadas a seguir antes de iniciar o desafio.

Atividades

Os alunos deverão desempenhar as seguintes atividades:

Obs.: usar o Google Colab.

Preparação do Desafio:

1. Instalar e importar as bibliotecas necessárias para o projeto.
2. Selecionar interativamente o ativo a ser estudado e o período de análise (início e fim).
3. Fazer o download da base de dados.
4. Exibir uma amostra dos dados, podendo o usuário escolher interativamente se deseja usar o comando “print” ou “display”.

Biblioteca Numpy:

1. Importar o conjunto de dados em um array Numpy.
2. Calcular estatísticas descritivas básicas para os preços das ações (mínima, máxima, média, mediana e desvio padrão).
3. Identificar e tratar possíveis valores ausentes ou discrepantes no conjunto de dados.

Biblioteca Pandas:

1. Criar uma coluna que represente a variação percentual diária no preço das ações.
2. Agrupar e sumarizar os dados por mês, calculando a média e a mediana das variações percentuais.
3. Exibir a média e a mediana da variação percentual mensal.

Biblioteca Matplotlib:

1. Plotar um gráfico de linha para visualizar a evolução do preço das ações ao longo do tempo.
2. Criar um gráfico de barras para mostrar o volume de negociações por mês.
3. Utilizar gráficos de dispersão para identificar possíveis correlações entre diferentes variáveis, como preço e volume.

Biblioteca Scikit-learn:

1º Modelo: Regressão (Previsão Preço)

1. Dividir o conjunto de dados em conjuntos de treinamento e teste.
2. Implementar um modelo de regressão para prever o preço das ações no próximo mês.
3. Avaliar o desempenho do modelo usando métricas como MSE (Mean Squared Error) e R2.
4. Plotar o resultado (Real x Previsto) .

2º Modelo: Classificação (Previsão Movimento do ativo)

Definindo o problema: vamos tentar prever se o preço das ações subirá ou descerá no dia seguinte, com base nas informações dos dias anteriores.

1. Criar um rótulo binário: 1 se o preço subiu e 0 se o preço desceu.
2. Usar características como preços e volumes anteriores.
3. Dividir o conjunto de dados em treinamento e teste.

4. Treinar um modelo de classificação.
5. Avaliar o modelo no conjunto de testes.
6. Visualizar os resultados.

Premissas:

- 1) Toda a codificação deverá ser muito bem comentada para facilitar análises e estudos futuros tanto do criador quanto de outros cientista de dados que porventura utilizem o código.
- 2) Para a seleção do ativo a ser estudado e o período de análise, deve ser feito via interação com o usuário (uso do “input”), facilitando assim usar o modelo para diversos ativos e diversos períodos.

Nota:

Para este desafio, você pode buscar conjuntos de dados de mercado financeiro em fontes públicas como Yahoo Finance, Google Finance, entre outros. Recomenda-se escolher dados de uma empresa conhecida e que tenha uma boa quantidade de dados históricos disponíveis para análise.

Sugestão de Fonte, Ativo e Período usado no modelo:

Fonte: Yahoo Finance

Ativo: BBDC4.SA (Banco Bradesco)

Período: 01/01/2008 a 01/01/2023 (15 anos)