

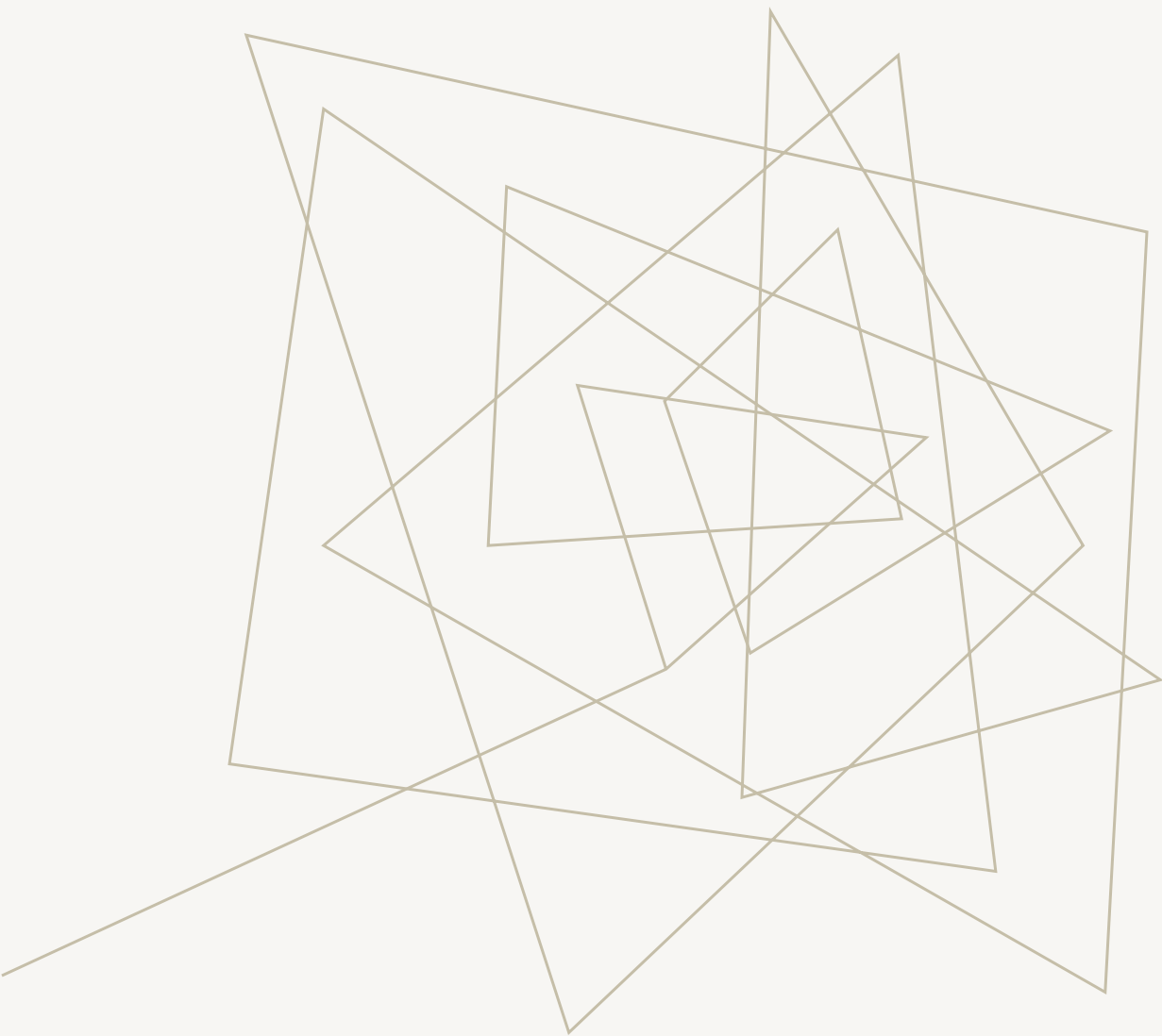


PPGI PROGRAMA
DE PÓS-GRADUAÇÃO
EM INFORMÁTICA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Fundamentos de Ciência de Dados

PPGI/UFRJ 2024.2
Prof. Sergio Serra &
Prof. Jorge Zavaleta

por: Cristiano FUSCHILO



Proposta

ANÁLISE DO MERCADO DE AÇÕES

Dados de séries temporais estão em todos os lugares, então analisá-los é importante para qualquer analista de dados ou cientista de dados.

DESAFIOS

- Descobrir e explorar dados do mercado de ações - ações de tecnologia (apple, amazon, google e microsoft).
- Usar o yfinance para obter informações sobre ações e visualizar diferentes aspectos delas usando seaborn e matplotlib.
- Analisar o risco de uma ação, com base em seu histórico de desempenho anterior.
- Prever preços futuros de ações por meio de um método de long short term memory (lstm).

PERGUNTAS QUE QUEREMOS RESPONDER

1. Qual foi a mudança no preço da ação ao longo do tempo?
2. Qual foi a média móvel das várias ações?
3. Qual foi o retorno diário médio da ação?
4. Qual foi a correlação entre as diferentes ações?
5. Quanto valor colocamos em risco ao investir em uma ação específica?
6. Como podemos tentar prever o comportamento futuro das ações?
(Prevendo o preço de fechamento das ações da APPLE inc usando LSTM)

OBTENDO OS DADOS

- O primeiro passo é obter os dados e carregá-los na memória.
 - Ações do site Yahoo Finance => rico recurso de dados do mercado financeiro e ferramentas para encontrar investimentos atraentes.
 - Biblioteca yfinance => baixa dados de mercado do Yahoo.
 - Biblioteca Pandas => trabalha com dados relacionais de maneira simples e intuitiva.

REPRODUTIVIDADE

- "A replicabilidade é uma pedra angular da ciência. Se um resultado experimental não pode ser obtido novamente por uma parte independente, ele se torna, na melhor das hipóteses, apenas uma observação que pode inspirar pesquisas futuras."

Mesirov

- É, ou deveria ser, a obtenção dos mesmos resultados em um estudo independente cujos procedimentos são os mais próximos possíveis do experimento original.

PROVENIÊNCIA

- “A proveniência é definida como um registro que descreve as pessoas, instituições, entidades e atividades envolvidas na produção, influência ou entrega de um dado ou coisa”.

Moreau e Groth

- Permite
 - Rastrear a origem de seus resultados
 - Repetir seus experimentos
 - Validar os processos que foram usados para derivar os dados

REPRODUTIVIDADE & PROVENIÊNCIA NESSE ESTUDO

- Reprodutividade
 - Uso do GitHub
 - Apresentação
 - Código fonte
 - Yahoo Finance
- Proveniência
 - Código fonte comentado explicando cada etapa do desenvolvimento

RESPONDENDO A PERGUNTA 1!

1. Qual foi a mudança no preço das ações ao longo do tempo?

- Usar a biblioteca Pandas para solicitação de informações de ações e análise dos atributos básicos de uma ação.
- Revisar o conteúdo dos dados
 - dados são numéricos e
 - a data é o índice dos dados.
 - os fins de semana estão faltando nos registros.

ESTATÍSTICAS DESCRITIVAS SOBRE OS DADOS

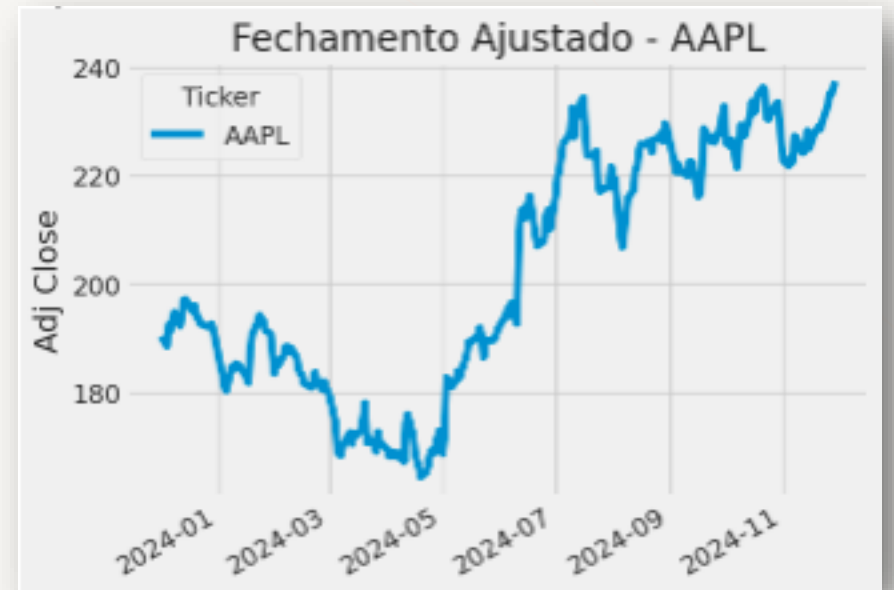
- `.describe()` => Gera Estatísticas Descritivas
 - resume a tendência central, dispersão e forma da distribuição de um conjunto de dados, excluindo valores NaN.
- Analisa séries numéricas e de objetos, bem como conjuntos de colunas DataFrame de tipos de dados mistos.
- Temos apenas 255 registros em um ano porque os fins de semana não estão incluídos nos dados.

INFORMAÇÕES SOBRE OS DADOS

- `.info()` => Imprime Informações
 - Método sobre um DataFrame, incluindo o dtype e as colunas do índice, valores não nulos e uso de memória.

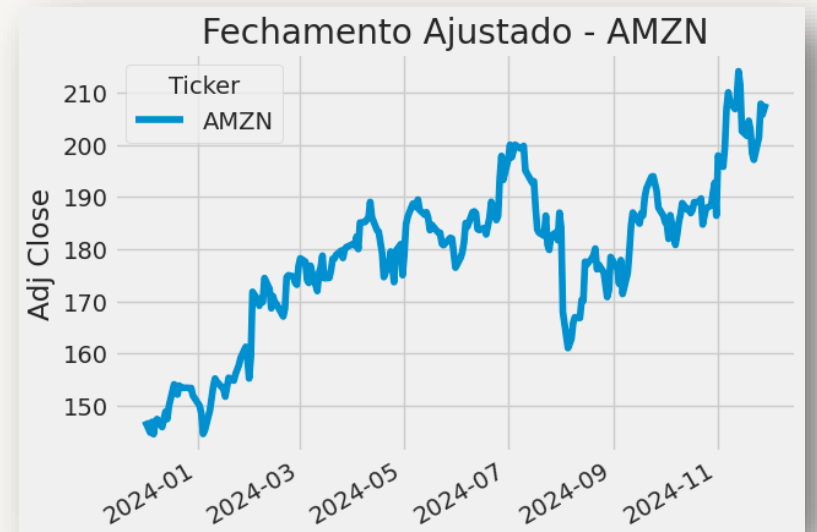
PREÇO DE FECHAMENTO

- Último preço pelo qual a ação é negociada durante o dia de negociação regular.
- Benchmark padrão usado por investidores para rastrear seu desempenho ao longo do tempo



VOLUME DE VENDAS

- Volume é a quantidade de um ativo ou título que muda de mãos ao longo de um período de tempo, geralmente ao longo de um dia.
- Por exemplo, o volume de negociação de ações se referiria ao número de ações de título negociadas entre sua abertura e fechamento diários.
- O volume de negociação e as mudanças no volume ao longo do tempo são entradas importantes para **traders** técnicos.



RESPONDENDO A PERGUNTA 2!

2. Qual foi a média móvel das várias ações?

- A média móvel (MA) é uma ferramenta simples de análise técnica que suaviza os dados de preço criando um preço médio constantemente atualizado.
- A média é tomada em um período de tempo específico, como 10 dias, 20 minutos, 30 semanas ou qualquer período de tempo que o *trader* escolher.

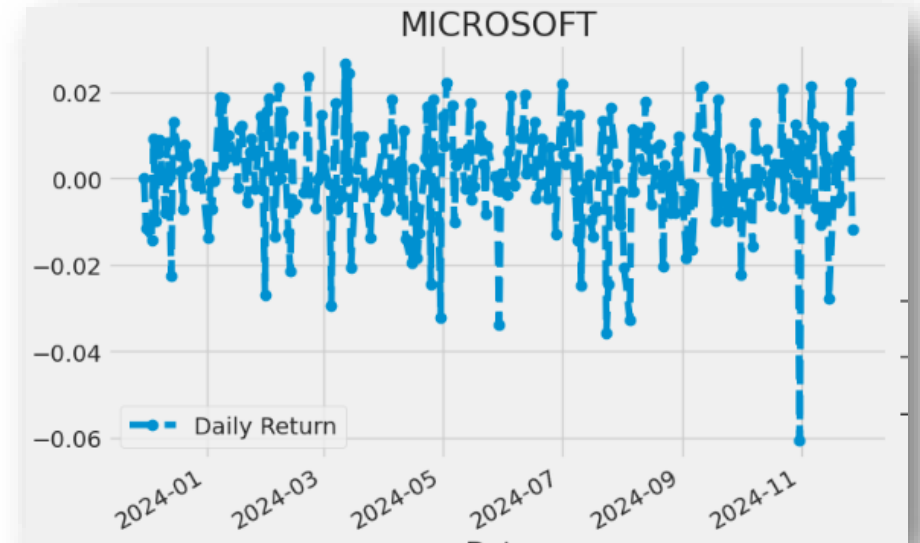
IPC: Os melhores valores para medir a média móvel são 10 e 20 dias porque ainda capturamos tendências nos dados sem ruído.



RESPONDENDO A PERGUNTA 3!

3. Qual foi o retorno diário médio da ação?

- Analisar o risco da ação.
 - Analisar => mudanças diárias da ação, e não apenas em seu valor absoluto.



RESPONDENDO A PERGUNTA 4!

4. Qual foi a correlação entre os preços de fechamento de diferentes ações?
- Correlação é uma estatística que mede o grau em que duas variáveis se movem em relação uma à outra, que tem um valor que deve estar entre -1,0 e +1,0. A correlação mede a associação, mas não mostra se x causa y ou vice-versa — ou se a associação é causada por um terceiro fator.
 - Analisar os retornos de todas as ações em nossa lista

CORRELAÇÃO ENTRE AS AÇÕES

- Então agora podemos ver que se duas ações são perfeitamente (e positivamente) correlacionadas entre si, uma relação linear entre seus valores de retorno diário deve ocorrer.
- Seaborn e pandas tornam muito fácil repetir essa análise de comparação para cada combinação possível de ações em nossa lista de ações.

ANALISE DA CORRELAÇÃO

- O gráfico plotado mostra todas as relações em retornos diários entre todas as ações.
Analisando temos:
 - Uma correlação interessante entre os retornos diários do Google e da Amazon. Pode ser interessante investigar essa comparação individual.
- Ferramentas:
 - `sns.pairplot()` => simples e direta.
 - `sns.PairGrid()` => controle total da figura (tipo de plotagens vão na diagonal, no triângulo superior e no triângulo inferior).

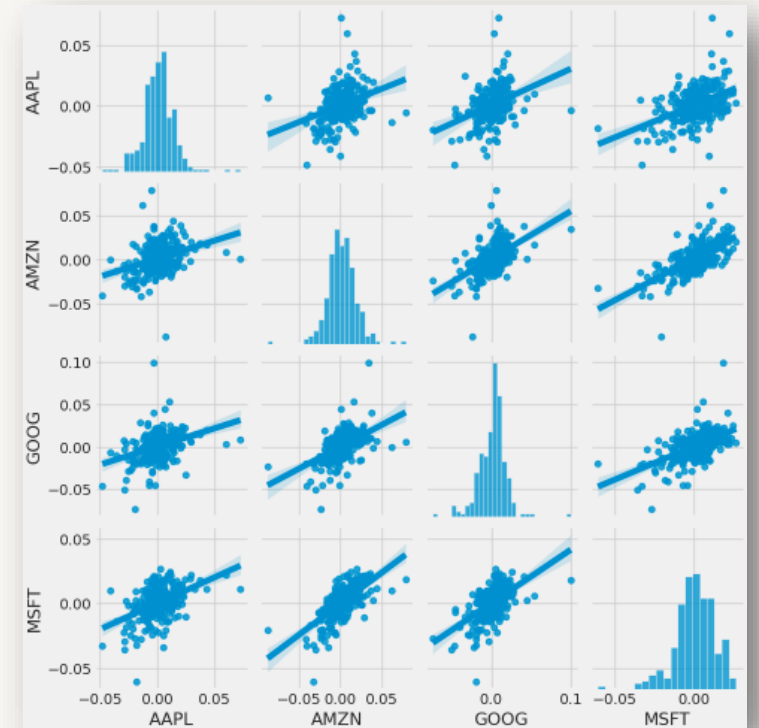
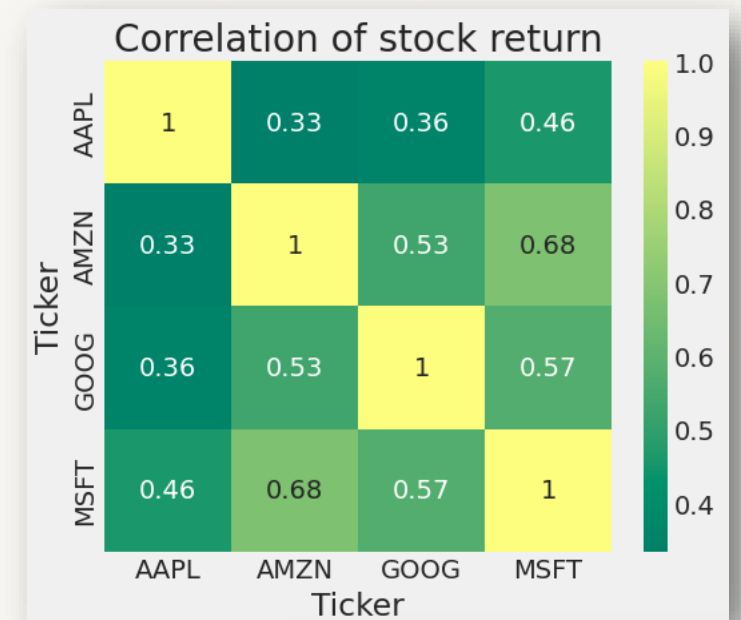


GRÁFICO DE CORRELAÇÃO

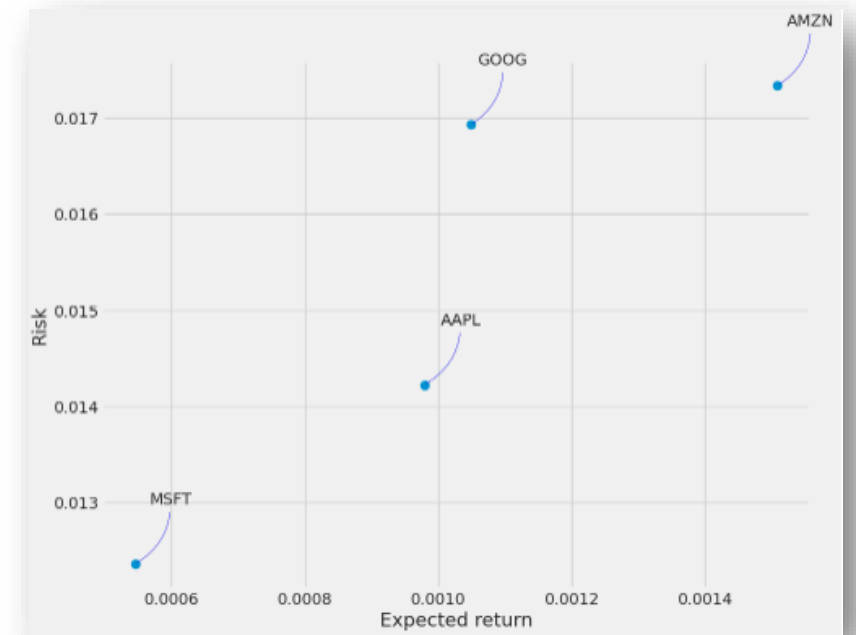
- Obter valores numéricos reais para a correlação entre os valores de retorno diário das ações.
- Ao comparar os preços de fechamento, vemos uma relação interessante entre a Microsoft e a Apple
- PairPlot => visualmente a Microsoft e a Amazon tiveram a correlação mais forte de retorno diário de ações.
- Também é interessante ver que todas as empresas de tecnologia estão positivamente correlacionadas.



RESPONDENDO A PERGUNTA 5!

5. Quanto valor colocamos em risco ao investir em uma ação específica?

- Existem muitas maneiras de quantificar o risco, uma das maneiras mais básicas de usar as informações que coletamos sobre retornos percentuais diários é comparando o retorno esperado com o desvio padrão dos retornos diários.



RESPONDENDO A PERGUNTA 6!

6. Previsão do preço de fechamento das ações da APPLE inc:

- A fim de responder essa pergunta é necessário seguir uma série de passos descritos a seguir!

A. OBTENDO OS DADOS

- Obter a cotação de ações
- Criar um novo dataframe com apenas a coluna preço de fechamento 'Close'
- Converter o dataframe em uma matriz numpy
- Obter o número de linhas para treinar o modelo

B. TREINAMENTO

- Escalar os dados
- Criar o conjunto de dados de treinamento
- Criar o conjunto de dados de treinamento escalado
- Dividir os dados em conjuntos de dados x_{train} e y_{train}
- Converter o x_{train} e y_{train} em matrizes numpy
- Remodelar os dados

C. MODELO LSTM

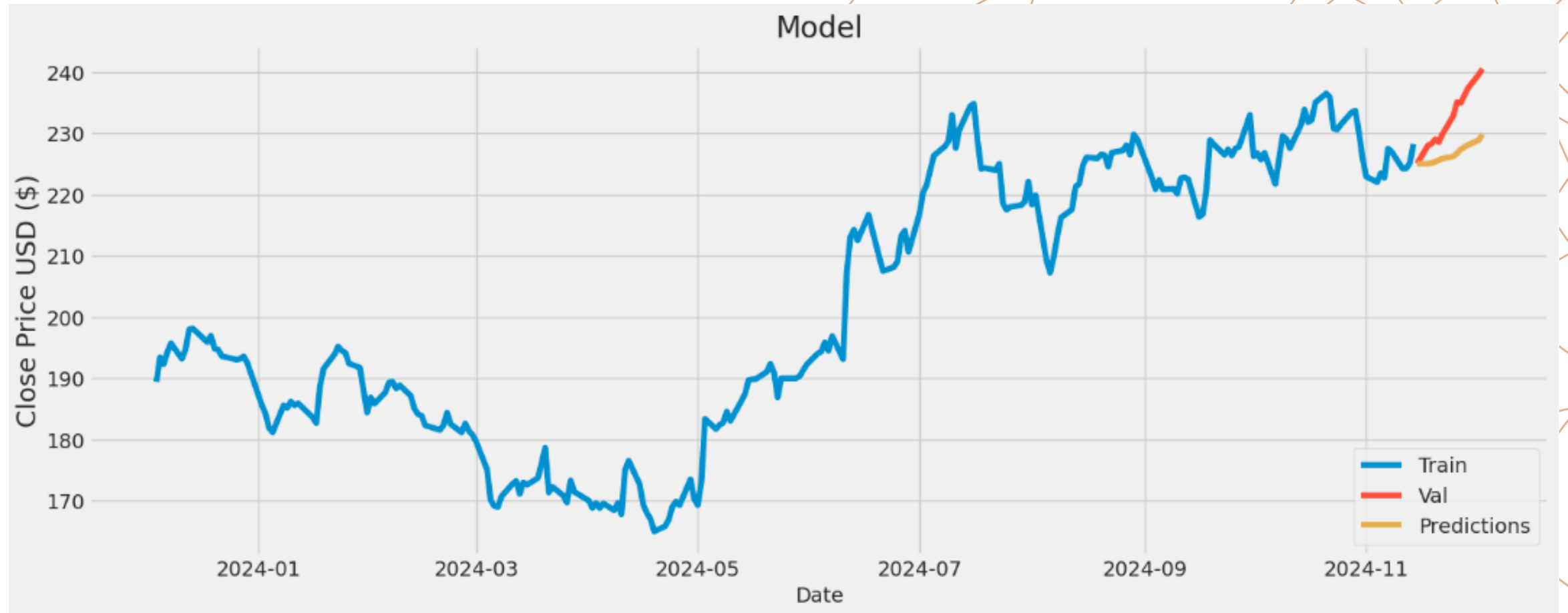
- Arquitetura de rede neural recorrente (RNN) que “*lembra*” valores em intervalos arbitrários.
- Adequada para classificar, processar e prever séries temporais com intervalos de tempo de duração desconhecida. A insensibilidade relativa ao comprimento do gap dá uma vantagem à LSTM em relação a RNNs tradicionais (também chamadas “vanilla”).

A estrutura de uma RNN é muito semelhante ao Modelo Oculto de Markov (modelo estatístico que assume que o estado de um sistema é parcialmente observável ou observável com ruído). No entanto, a principal diferença é como os parâmetros são calculados e construídos.

C. MODELO LSTM

- Construir o modelo LSTM
- Compilar o modelo
- Treinar o modelo
- Criar o conjunto de dados de teste
- Criar uma nova matriz contendo valores escalados

RESULTADO



TRABALHOS FUTUROS

- Aprimorar a Proveniência
- Usar “papéis” brasileiros
- Fazer treinamentos repetitivos (pipeline)
- Trabalhar com valores de log retorno da ação ao invés do preço
- Usar outras heurísticas de treinamento

Abstract geometric lines in the top left corner, consisting of several thin, light brown lines that intersect to form various polygons and shapes.

OBRIGADO

Cristiano FUSCHILO

21 99242-0699

Fuschilo@gmail.com