

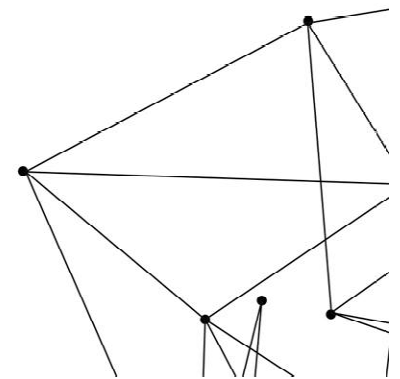


# **REGRESSÃO EM SÉRIES TEMPORAIS FINANCEIRAS COM RNN: UM ESTUDO COM MILHO FUTURO**

Gustavo Trielli Avila

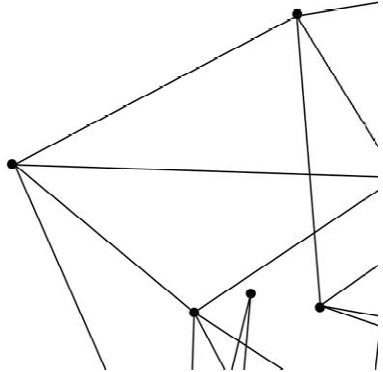
Orientador: Dr. Clayton Reginaldo Pereira

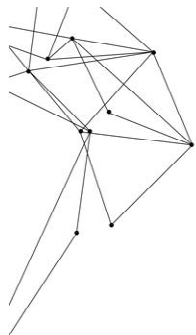
**UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"**  
FACULDADE DE CIÊNCIAS - CAMPUS BAURU  
DEPARTAMENTO DE COMPUTAÇÃO  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO





# CONTEÚDO

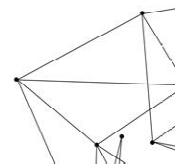
- 01 Introdução
  - 02 Metodologia
  - 03 Desenvolvimento
  - 04 Resultados
  - 05 Conclusão
- 

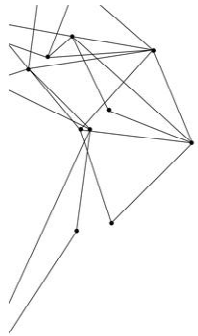


# Introdução

## A Expansão do Mercado Financeiro

- Aumento de 133% no número de CPFs cadastrados na Bolsa nos últimos dois anos.
- Investidores buscam estratégias que tenham maior retorno e menor risco.

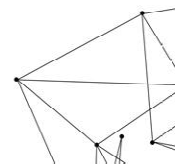


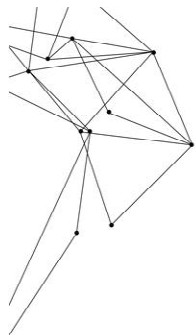


# Introdução

## Predição em séries temporais financeiras

- Análise fundamentalista.
- Análise técnica.
- Utilização de Redes Neurais.

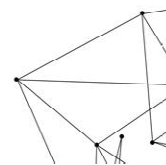




# Introdução

## O Milho Futuro

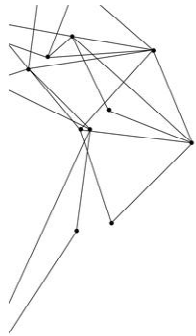
- Derivativo agrícola.
- Menor exposição à influências externas no preço.
- Comportamento sazonal devido a fatores climáticos.
- Maior retorno ao investidor.





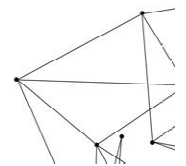
# Metodologia

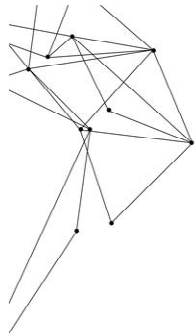




# Indicadores técnicos

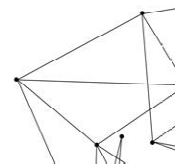
- Média Móvel
- Média Móvel Exponencial
- Histograma MACD
- Índice de força relativa (RSI)





## Séries exógenas

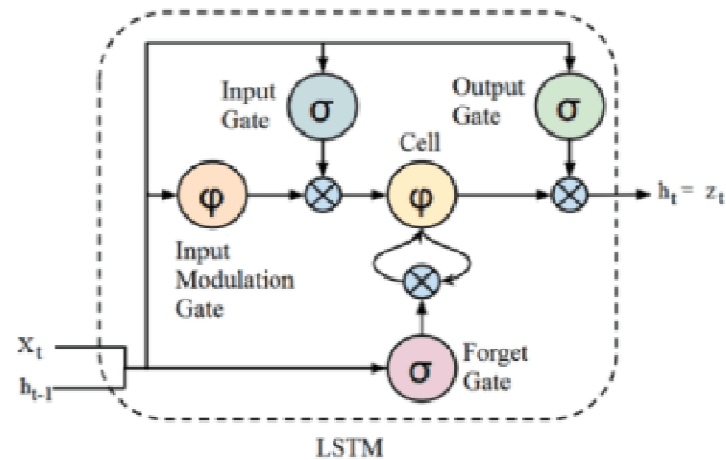
- Indicador de preços do milho Esalq/BM&FBOVESPA
- Série de histórica do índice DI

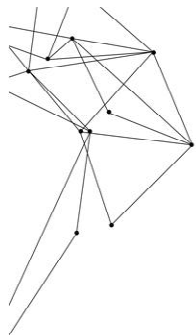




# LSTM

- Tipo de Rede Neural Recorrente que consegue armazenar dependências temporais de longo prazo





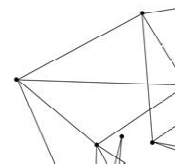
# Modelos Propostos

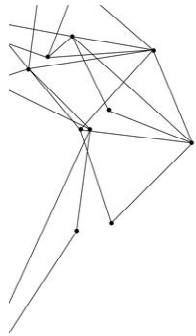
## Inputs

- **Conjunto 1:** Somente dados de negociação (Abertura, Máxima, Mínima, Fechamento, Volume)
- **Conjunto 2:** Dados de negociação e indicadores técnicos
- **Conjunto 3:** Dados de negociação, indicadores técnicos, e séries exógenas

## Tamanho da sequência

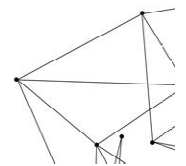
- **2 dias**
- **5 dias**
- **15 dias**





# Análise dos Dados

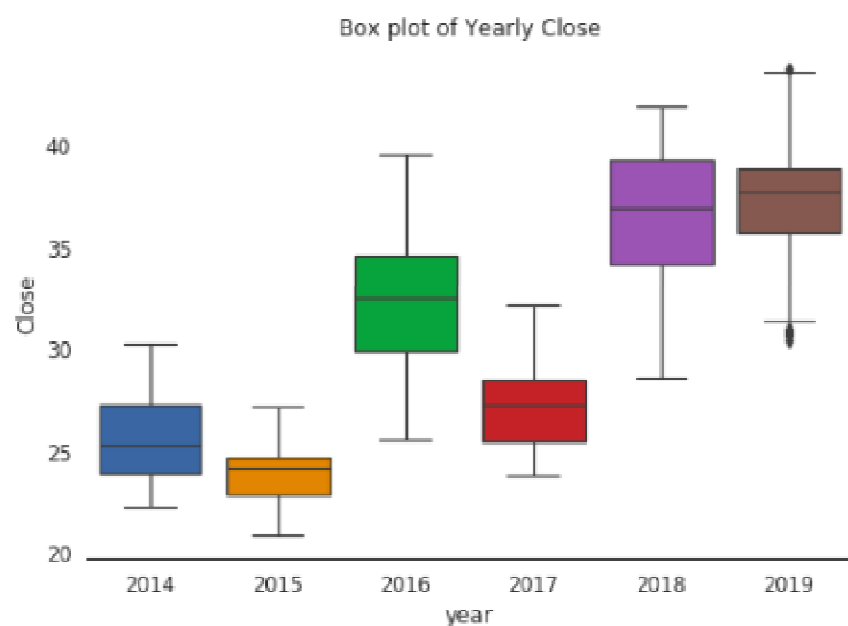
	Open	High	Low	Close	Volume
Date					
2014-05-15	28.66	28.86	28.42	28.57	3152
2014-05-16	28.63	28.76	28.52	28.55	1915
2014-05-19	28.42	28.44	28.21	28.21	2867
2014-05-20	28.28	28.30	27.96	27.97	2111
2014-05-21	27.89	27.90	27.71	27.81	2515
2014-05-22	27.90	28.16	27.90	28.06	1681
2014-05-23	28.01	28.17	27.93	27.99	1662
2014-05-26	28.11	28.42	28.02	28.15	866
2014-05-27	27.85	27.89	27.72	27.80	2508
2014-05-28	27.80	27.86	27.70	27.78	2571



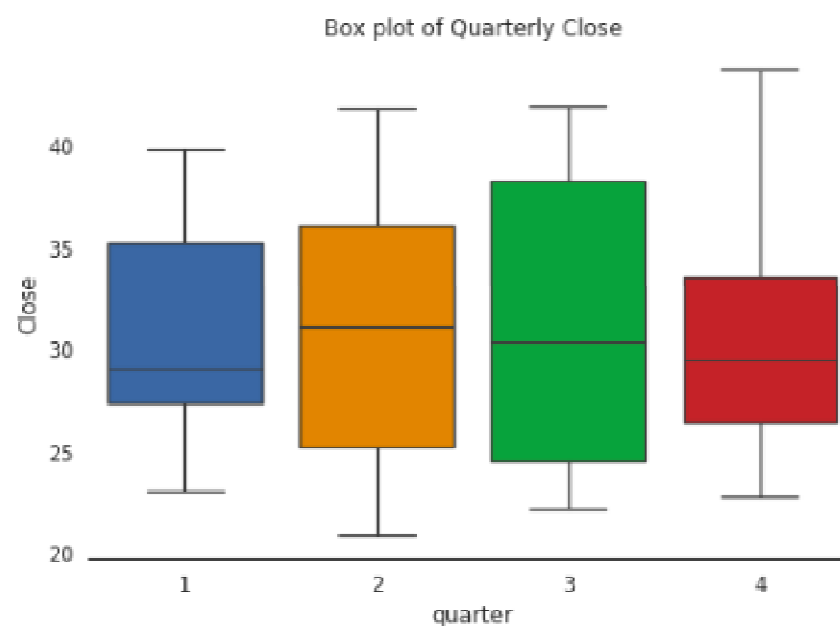
# Análise dos Dados

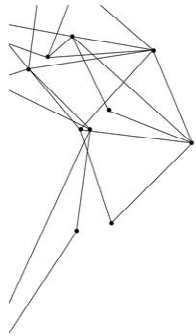


# Análise dos Dados



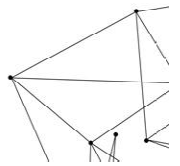
Box plot of Quarterly Volume

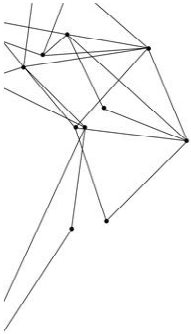




# Pré-processamento

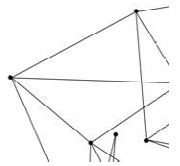
- Manipulação para inserção das novas variáveis
- Normalização dos dados
- Separação em conjunto de treino, validação e teste.





# Treinamento

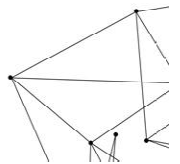
- Os modelos possuem 3 camadas, a camada de input, a camada LSTM bidirecional, e a camada de saída
- A entrada possui dimensão de  $(n \times t)$ , onde  $n$  é o tamanho da sequência e  $t$  é o número de atributos. A camada LSTM possui dimensão de  $(2 \times t)$ . A camada de saída possui 1 dimensão
- Função *tanh* como função de ativação tanto na camada de saída quanto na camada LSTM
- 2000 épocas de treinamento ou até que parasse de convergir



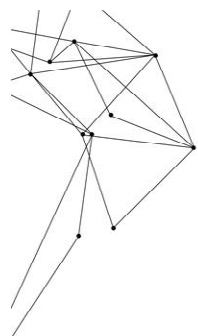


# Métricas de desempenho

- Erro percentual absoluto médio (MAPE)
- Raiz quadrada do erro quadrático médio (RMSE)
- Acurácia da previsão da direção em relação ao último fechamento (Acc1)
- Acurácia da previsão da direção em relação à última previsão (Acc2)



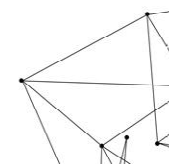


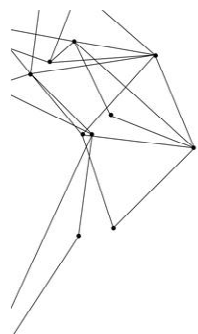


# Análise dos Resultados

## Treino

conjunto 1	2 dias 1,2910%	5 dias 1,6649%	15 dias 1,5852%	MAPE
conjunto 2	1,2985%	1,2070%	1,5056%	MAPE
conjunto 3	1,7956%	<b>1,1406%</b>	1,4470%	MAPE
conjunto 1	0,4973	0,6090	0,5864	RMSE
conjunto 2	0,4892	0,4547	0,5515	RMSE
conjunto 3	0,6661	<b>0,4304</b>	0,5340	RMSE
conjunto 1	51,1401%	48,5342%	49,1857%	Acc1
conjunto 2	52,1173%	55,5917%	49,2942%	Acc1
conjunto 3	51,0315%	<b>59,6091%</b>	53,6374%	Acc1
conjunto 1	56,3043%	52,6087%	55,6522%	Acc2
conjunto 2	55,8696%	<b>56,9565%</b>	55,9783%	Acc2
conjunto 3	54,2391%	56,8478%	56,4130%	Acc2

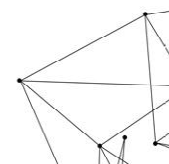


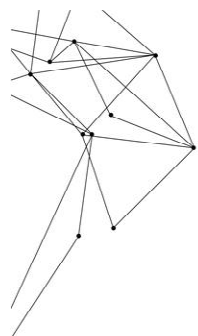


# Análise dos Resultados

## Validação

	2 dias	5 dias	15 dias	
conjunto 1	1,4362%	1,2150%	<b>1,1988%</b>	MAPE
conjunto 2	1,4495%	1,3669%	1,8553%	MAPE
conjunto 3	1,7582%	2,2063%	2,0011%	MAPE
conjunto 1	0,6952	<b>0,5705</b>	0,5927	RMSE
conjunto 2	0,7284	0,6753	0,8916	RMSE
conjunto 3	0,8370	1,0364	0,9796	RMSE
conjunto 1	49,8099%	49,0494%	51,7110%	Acc1
conjunto 2	50,1901%	48,2890%	47,5285%	Acc1
conjunto 3	48,2890%	<b>53,9924%</b>	50,9506%	Acc1
conjunto 1	53,8168%	56,1069%	57,6336%	Acc2
conjunto 2	55,8696%	54,5802%	<b>58,0153%</b>	Acc2
conjunto 3	50,3817%	55,7252%	54,9618%	Acc2

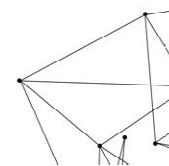


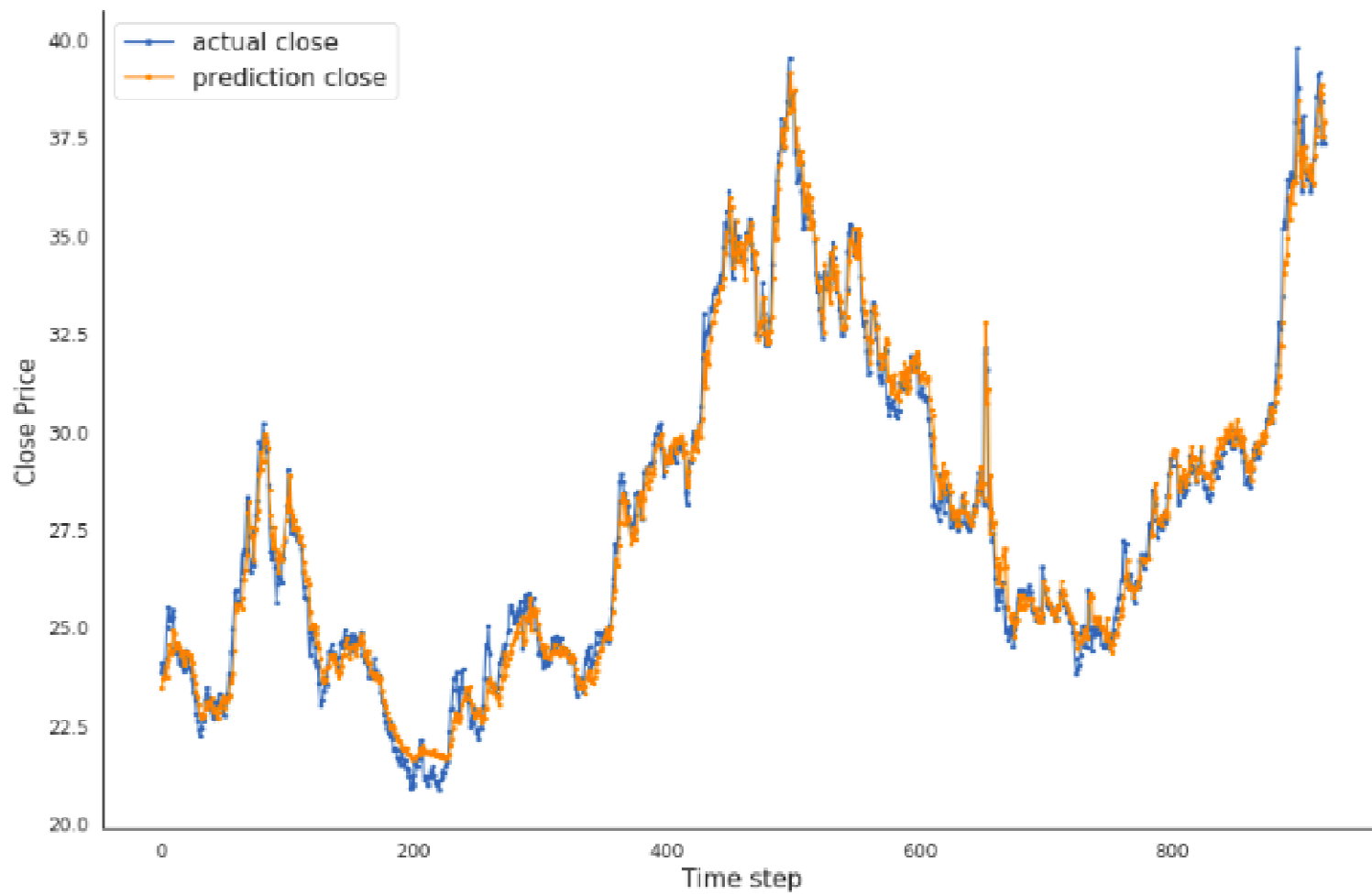


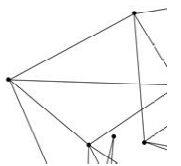
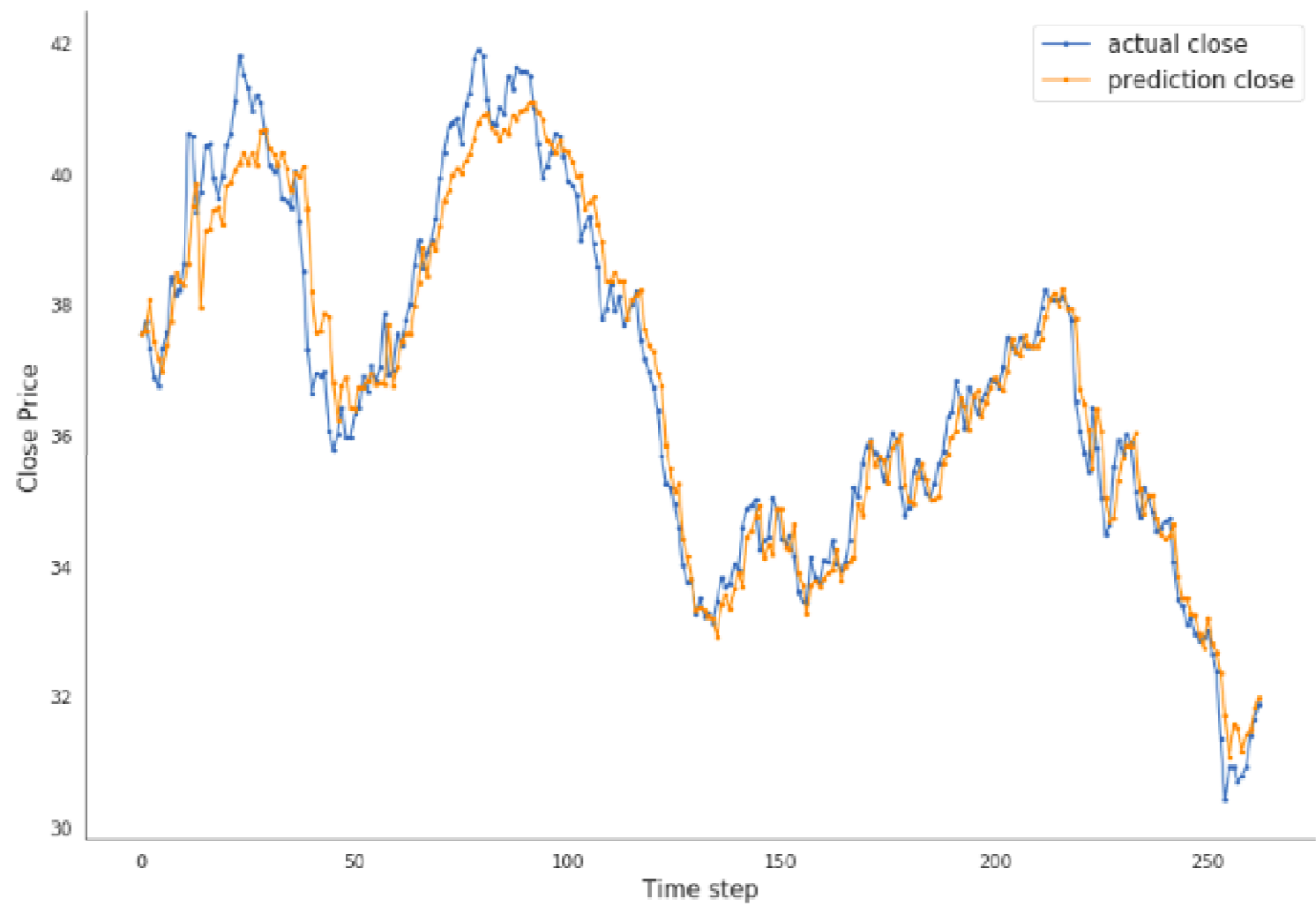
# Análise dos Resultados

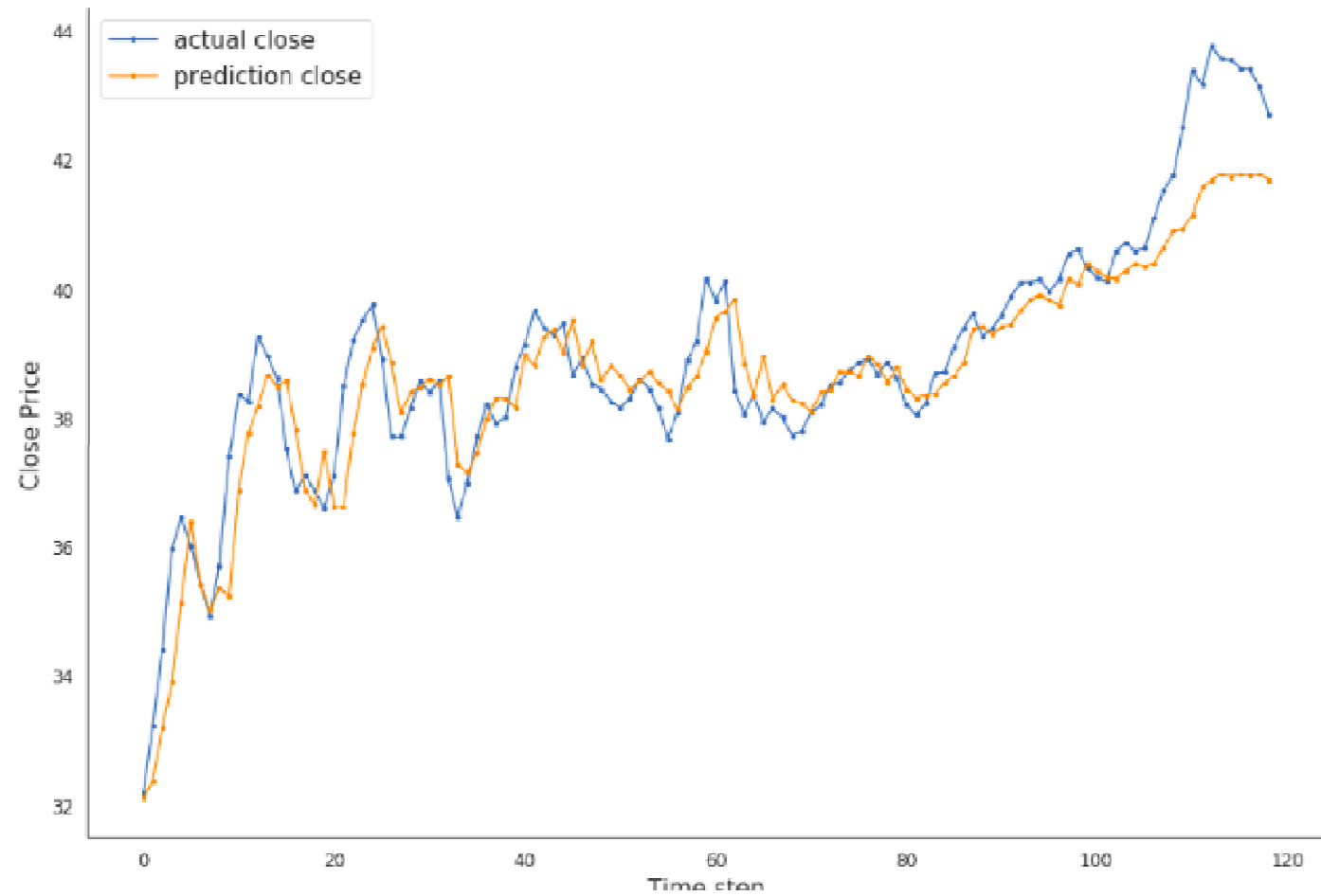
## Teste

	2 dias	5 dias	15 dias	
conjunto 1	1,6879%	1,6439%	<b>1,4800%</b>	<b>MAPE</b>
conjunto 2	1,8488%	1,7318%	2,1861%	<b>MAPE</b>
conjunto 3	2,1197%	2,0139%	2,1028%	<b>MAPE</b>
conjunto 1	0,9641	0,8559	<b>0,8035</b>	<b>RMSE</b>
conjunto 2	0,8711	0,8829	1,0112	<b>RMSE</b>
conjunto 3	1,0693	1,0046	1,0213	<b>RMSE</b>
conjunto 1	54,5455%	48,0620%	<b>54,6218%</b>	<b>Acc1</b>
conjunto 2	44,6970%	50,3876%	45,3782%	<b>Acc1</b>
conjunto 3	52,2727%	44,1860%	49,5798%	<b>Acc1</b>
conjunto 1	55,7252%	<b>60,1562%</b>	57,6271%	<b>Acc2</b>
conjunto 2	56,4885%	57,8125%	55,9322%	<b>Acc2</b>
conjunto 3	54,9618%	51,5625%	56,7727%	<b>Acc2</b>





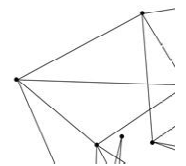






## Conclusão

- Existe uma dependência temporal em relação a série de preços e as redes LSTM podem ser úteis na previsão de séries temporais financeiras
- Inserção de variáveis além dos dados de negociação criam um modelo menos generalista
- Modelo de regressão não é a melhor abordagem para especulação, já que o erro não está diretamente ligado ao acerto na direção do movimento





# **OBRIGADO**

---

Gustavo Trielli Avila

Orientador: Dr. Clayton Reginaldo Pereira

