

Métodos de *Deep Learning* Aplicados a *Candlestick* como Estratégia de Investimento

Felipe Dias Paiva
Carolina M. Roma

Maio/2019



INTRODUÇÃO

- A previsão dos retornos das ações é considerada uma das tarefas mais desafiadoras do processo que lida com séries temporais financeiras, pois o mercado é essencialmente dinâmico, complexo, evolucionário, não linear, nebuloso, não paramétrico e caótico por natureza.
- Sistemas especialistas têm a sua utilização cada vez mais demandada, como resposta às características do ambiente, buscando identificar um nexo entre passado e futuro, com o objetivo de prever retorno ou o movimento de uma ação.
- As técnicas de inteligência artificial que permitem utilizar dados de diferentes naturezas e ainda não estão sujeitas à rigidez de pressupostos, como os que são impostos aos modelos econométricos.



INTRODUÇÃO

- O objetivo desta pesquisa foi investigar a capacidade dos métodos *Convolutional AutoEncoder* (CAE) e *Long Short-Term Memory* (LSTM) em prever o movimento do preço de ações negociadas no mercado acionário brasileiro.
- As principais contribuições desta pesquisa estão:
 - i) na investigação de métodos de *deep learning*, em específico o CAE e LSTM em conjunto.
 - i) fornecimento de novas evidências acerca do desempenho de métodos de *deep learning* para previsão dos preços com dados dados de um país emergente e não asiático.



CONVOLUTIONAL AUTOENCODER (CAE)

- A técnica CAE, método de *deep learning*, é utilizada para reduzir a dimensionalidade de imagens.
- A CAE é um tipo de Redes Neurais Convolucionais (CNNs): a CNN é treinada de ponta a ponta para aprender filtros e combinar recursos com o objetivo de classificar suas entradas. Já a CAE é treinada apenas para aprender filtros capazes de extrair recursos que podem ser usados para reconstruir a entrada.



CONVOLUTIONAL AUTOENCODER (CAE)

- As RNNs são um tipo de ANN que visa trabalhar com dados sequenciais.
- Versão mais robusta da RNN, a LSTM evita o problema de dependência de longo prazo. Manter na sua estrutura informações por longos períodos de tempo trata-se do seu comportamento padrão.
- As LSTMs possuem a estrutura de propagação em cadeia, além do módulo de repetição (loop. As RNNs padrão também possuem o módulo de repetição, porém numa estrutura mais simples, dado que operam com uma única camada.



DESIGN DO EXPERIMENTO

AMOSTRA: ações preferenciais da Petrobrás (PETR4).

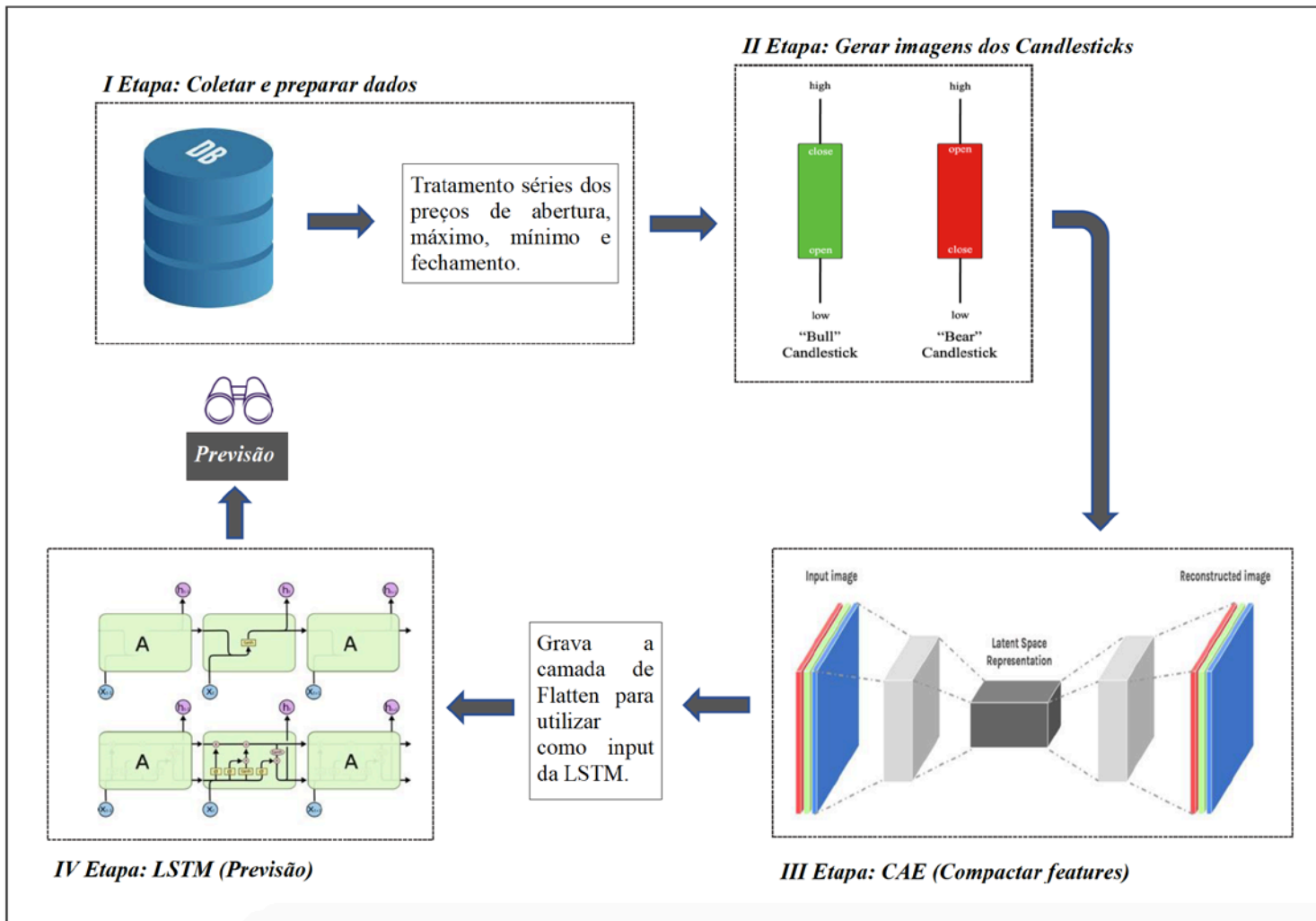
PERÍODO: 02/01/2001 a 29/09/2016.

COLETA DE DADOS: Terminal Bloomberg.

DATASET: Preços ajustados de Abertura, Máximo, Mínimo, Fechamento.



Esquema do experimento



Coleta e Preparação dos dados

- A primeira etapa foi calcular as variáveis que compõem o gráfico de Candlesticks Heikin-Ashi (HA). Os quatro pontos necessários para formação do HA são assim estabelecidos:

$$HA_Fech_t = (Abert_t + Max_t + Min_t + Fech_t) / 4$$

$$HA_Abert_t = (HA_Abert)_{t-1} + HA_Fech_t) / 2$$

$$HA_Max_t = \text{Maior} (Max_t, HA_Abert_t, HA_Fech_t)$$

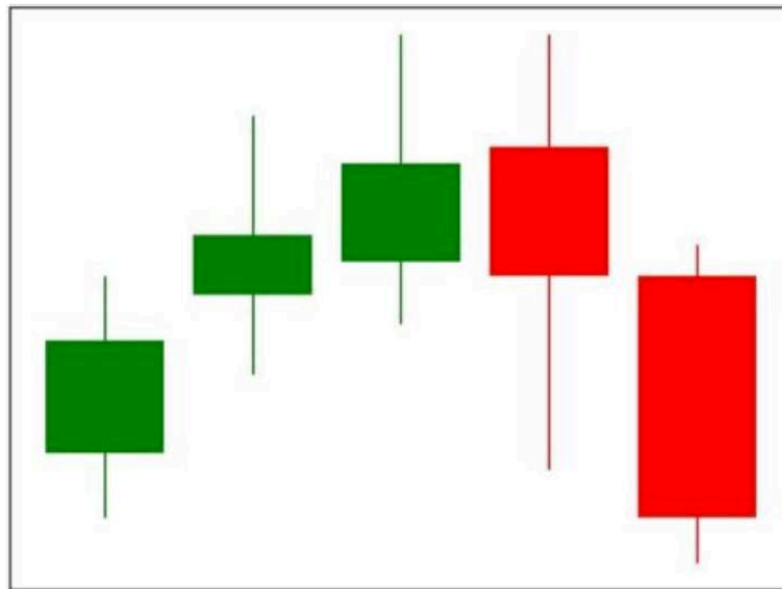
$$HA_Min_t = \text{Menor} (Min_t, HA_Abert_t, HA_Fech_t)$$

O gráfico HA visa facilitar a marcação das tendências de alta e de baixa.



Geração do Gráficos dos *Candlesticks* Heikin-Ashi

- Cada imagens do HA foi gerada contendo 5 dias de pregão.
- Foram geradas 3.900 imagens em formato de cor RGB, tamanho 128 x 128 pixels, ou seja, cada imagem possui 16.384 pontos para cada uma das escalas de cor.



Convolutational AutoEncoder dos Candlesticks

- Geradas e arquivadas as imagens dos HA, inicia-se a terceira etapa que foi gerar a *Convolutational AutoEncoder* (CAE).
- A etapa de treinamento da CAE com 3.800 imagens e o teste com 100 imagens.



Convolutational AutoEncoder dos Candlesticks

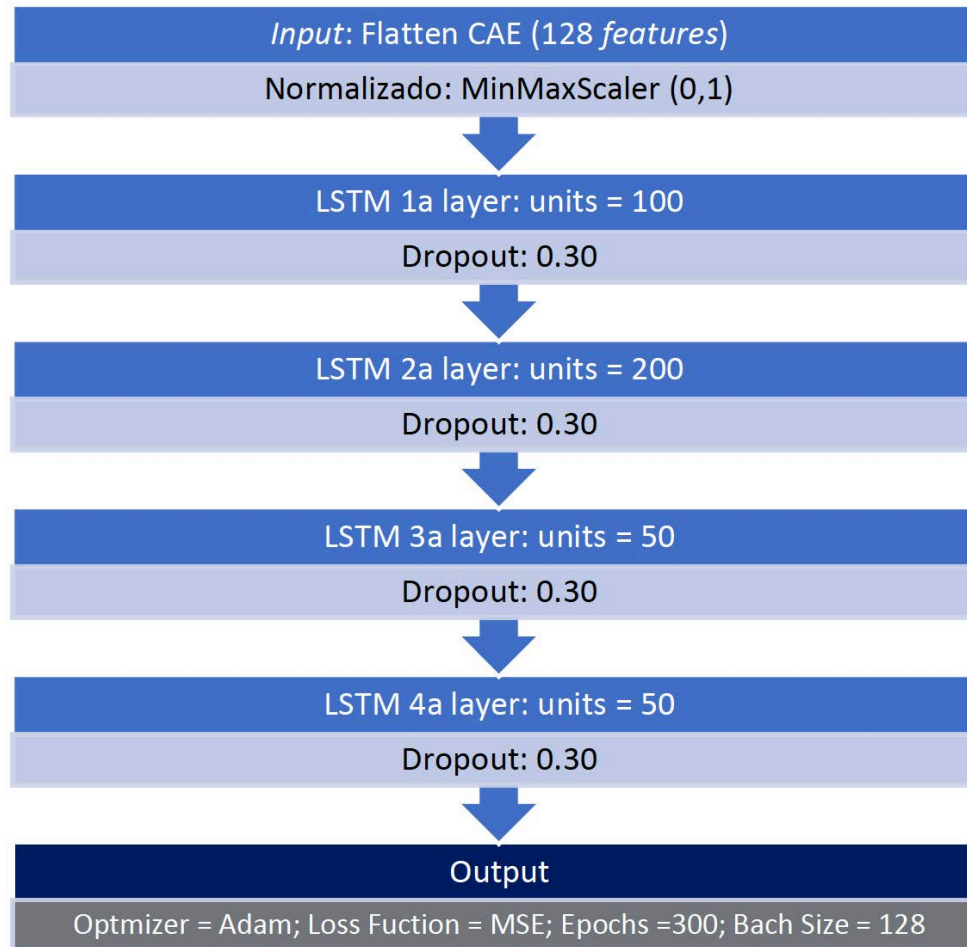


Long Short-Term Memory para Previsão dos Retornos

- Inputs: *features* geradas pelo CAE.
- Output: retorno máximo alcançado nos 3 dias subsequentes aos dados de input.
- Retorno máximo foi calculado estabelecendo que o investidor compra o ativo no preço de abertura do dia seguinte a formação da imagem.
- *Lag* de 90 imagens. Assim, para a fase de treinamento passou-se a ter 3.710 instâncias e 100 instâncias para o conjunto de teste.

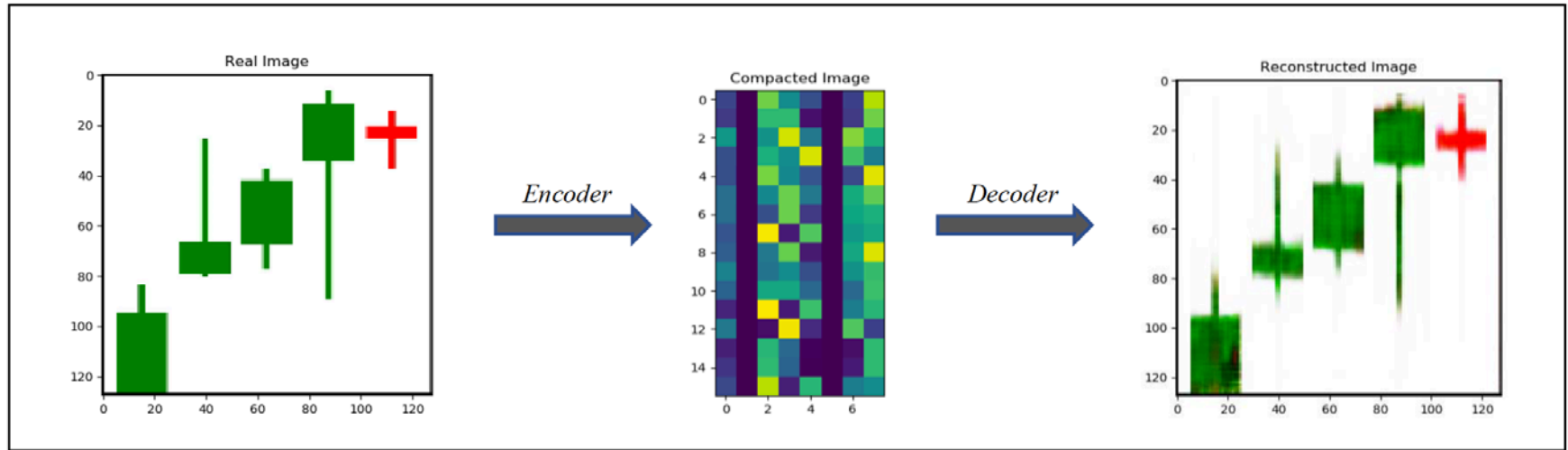


Long Short-Term Memory para Previsão dos Retornos



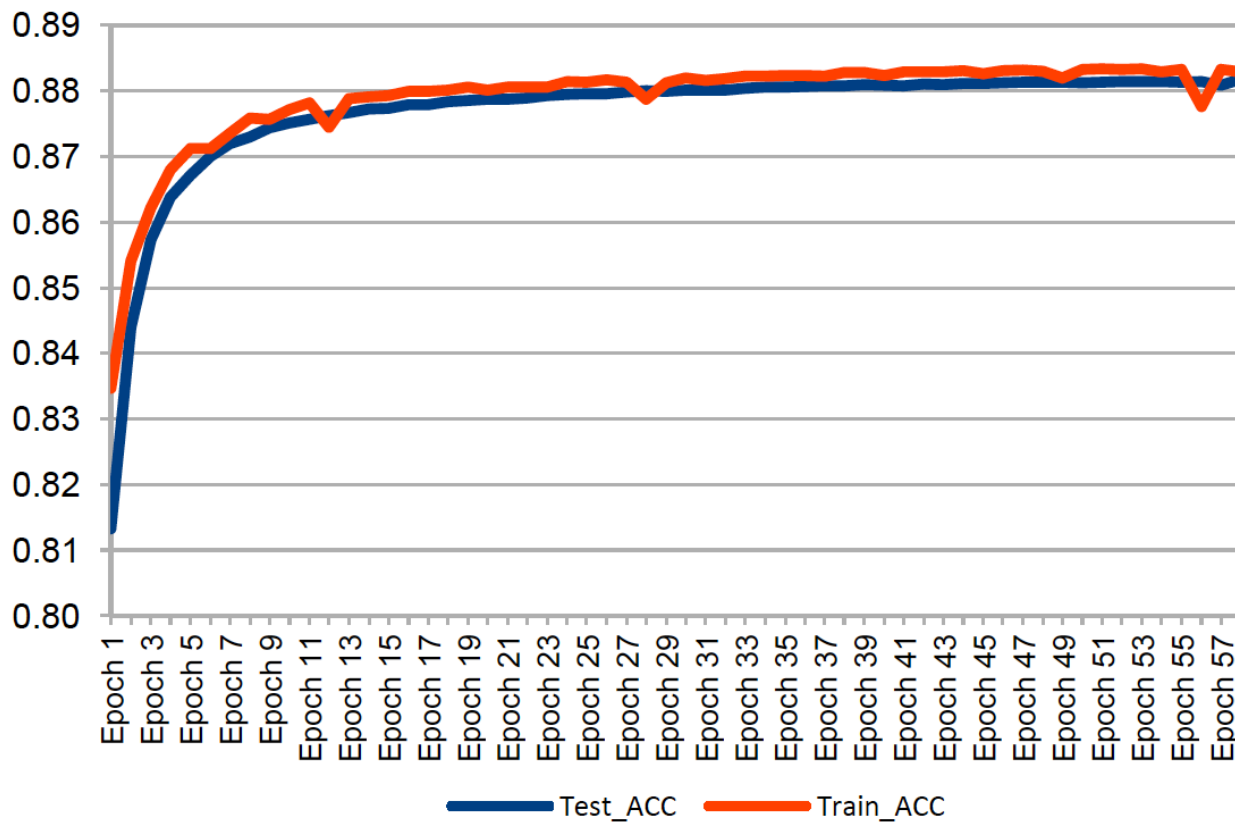
Resultados

- Exemplo do processo de encoder e decoder baseado na técnica de convolucional.



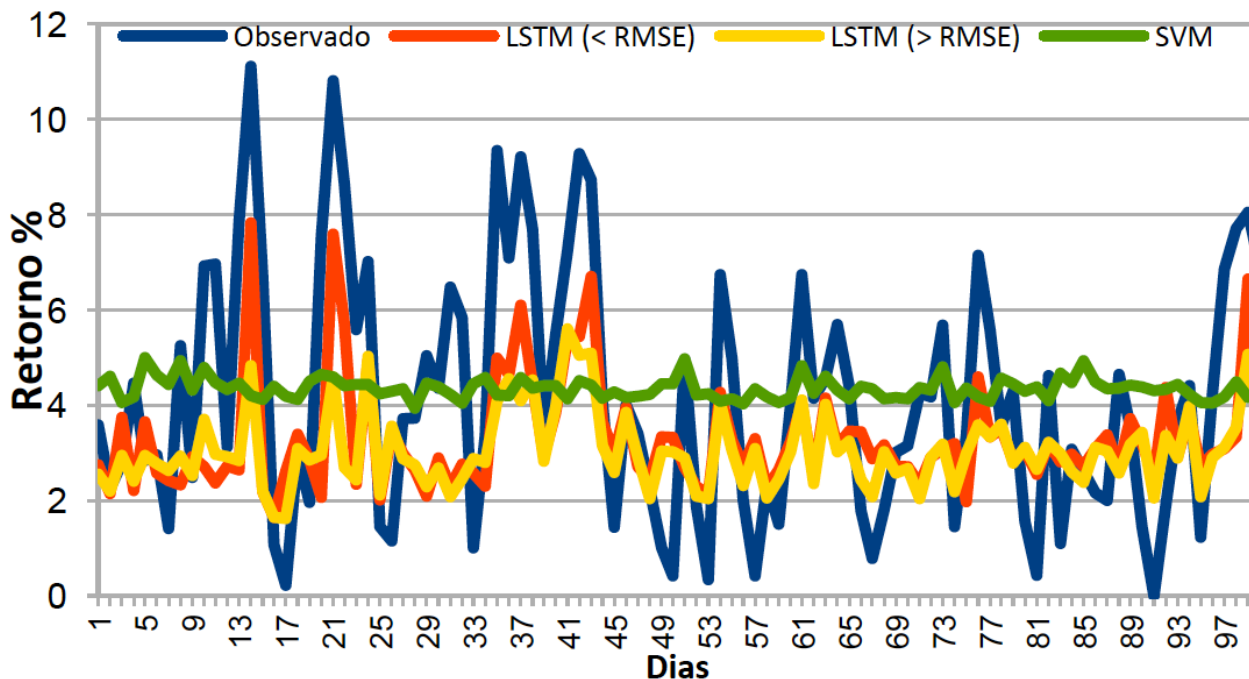
Resultados

- Acurácia média da imagens reconstruídas face a imagens reais foi de 88%, tendo convergido em 57 épocas.



Resultados

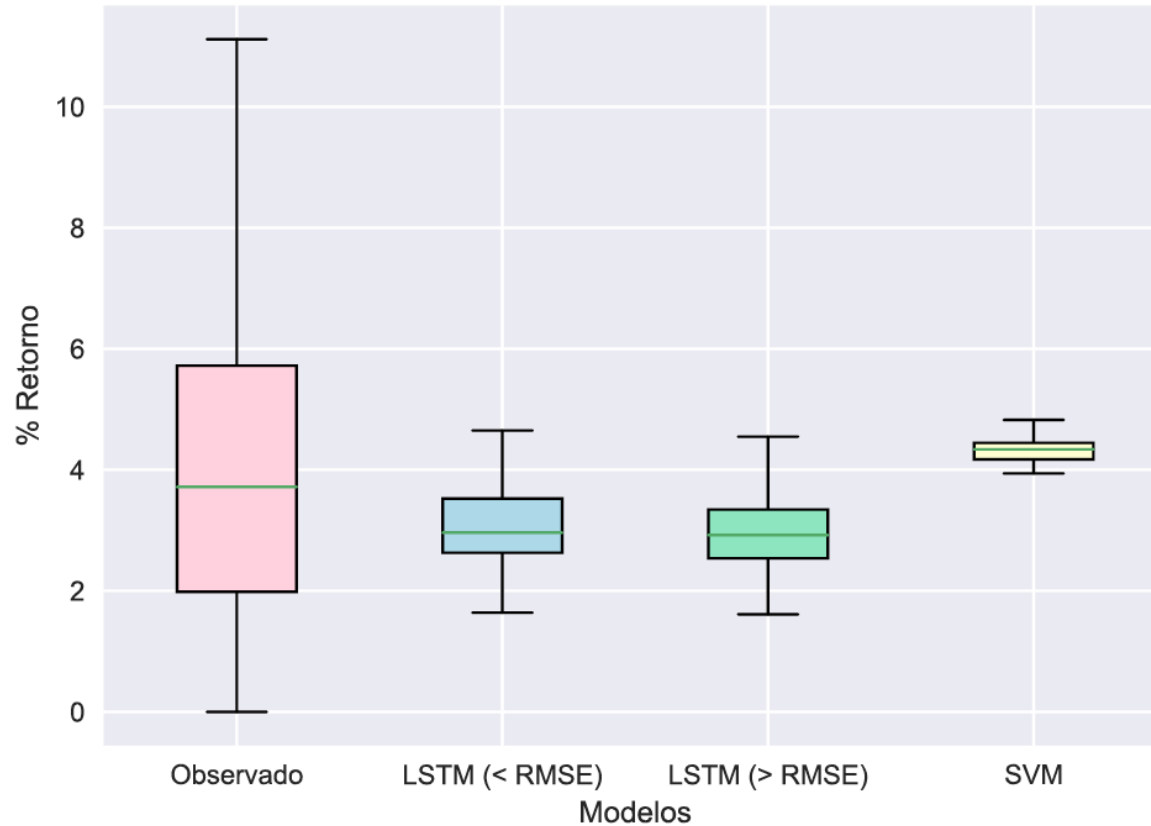
- Previsão dos retornos %:



- RMSE foi de 0,0257 para a SVM e de 0,02221 e 0,02404 para a LSTM de menor e maior RMSE entre as 50 iterações realizadas.

Resultados

- Box-Plot da distribuição da previsão do retorno %:



RESULTADOS

- Uma outra faceta para analisar os resultados num processo de tomadas de decisão de investimento, temos que caso fosse estabelecido um alvo de 3% de retorno por investimento, tem-se que das 50 iterações e que a decisão de investimento apenas seria tomada onde as situações de previsões superaram a maioria das previsões.
- Precision dos sinais de investimento de 73%. Sendo que, as decisões *naive* alcançam um *precision* de 62%.



CONSIDERAÇÕES FINAIS

- O modelo proposto apresentou resultado bastante interessante quando comparado aos *baselines*.
- As técnicas CAE e LSTM possuem um instigante e promissor potencial para auxiliar o investidor no processo de tomada de decisão. Tratam-se de técnicas que possuem uma ampla condição de exploração científica, com exigência de estudos amplos e consistentes.



TRABALHOS FUTUROS:

- Ampliar a base de dados, seja por questão temporal ou de outros ativos.
- Refinamento dos hiperparâmetros.
- Integração com outras técnicas de *machine learning*.



OBRIGADO!!!

Felipe Dias Paiva
fpaiva@cefetmg.br

