

# Forecasting the direction of High-Frequency Returns:

An ensemble-trees application

Rafael Silva Wagner  
André Alves Portela Santos

Federal University from Santa Catarina

14/08/2018

# Introdução

- ▶ Desde o trabalho de Christoffersen et al. (2006) e Christoffersen e Diebold (2006) prever os sinais dos retornos se tornou um ponto central em trabalhos de finanças aplicadas;
- ▶ Retornos intra-diários requerem uma abordagem diferente devido a sua difícil previsibilidade baseada em medidas estatísticas;
- ▶ Em janelas intra-diárias estimadores de micro-estrutura de mercado apresentam correlação significativa com a evolução dos preços (CAO; HANSCH; WANG, 2009);
- ▶ O alto volume de dados e a riqueza de estimadores da microestrutura se beneficiam das técnicas de estimação paralela e redução de dimensionalidade, características recorrentes em implantações de machine learning.

## Estimadores de Microestrutura de Mercado

- ▶ Queremos estimadores da dinâmica de negociação que nos ajudem a capturar movimentos com informações privilegiadas e persistências na dinâmica de negociação;
- ▶ Em sistemas complexos com alta presença de ruído é comum haver dificuldade em modelar e encontrar relações de causalidade; (TONG, 2013)
- ▶ O impacto teórico dessa característica é o pluralismo de trabalhos verificando fortes correlações estatísticas entre todos os pontos do sistema;
- ▶ Variáveis: Tempo; Último Preço, Retorno e Sinal; Volatilidade; Numero e Quantidade de Negócios; Volume e VWAP; Ordens novas, atualizadas e canceladas; Preço Maximo, Minimo e Médio das Ofertas.

# Aprendizado de Máquina

- ▶ Muitos trabalhos recentes vem verificando a capacidade de diferentes familias de algoritmos modelarem esse relacionamento entre a dinâmica da negociação e o preço dos ativos;
1. Redes Neurais - Dixon (2017) e Tsantekidis et al. (2017)
  2. SVM - Fletcher e Shawe-Taylor (2013) e Kercheval e Zhang (2015)
  3. Ensemble Trees - Han et al. (2015)
- ▶ Devido ao recente sucesso de modelos de árvores no Kaggle nós nos interessamos em estender o trabalho de Han et al. (2015) que apenas avaliava o algoritmo Random forests para dois outros algoritmos.

# Árvores de Decisão

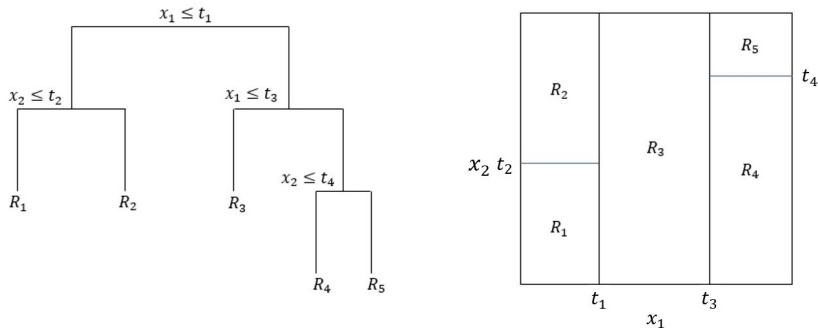


Figure 1: Example of Decision Tree

## Pontos positivos de Modelos de Conjuntos de Árvores

- ▶ Processo de estimação simplificado e altamente paralelizado;
- ▶ Seleção de variáveis integrada a estimação;
- ▶ Robustez a ruídos;
- ▶ Alto poder preditivo.

# Conjuntos de Árvores de Decisão

Modelos avaliados por nós:

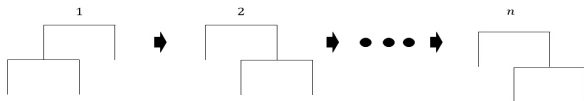


Figure 2: Boosting Trees - Chen e Guestrin (2016)

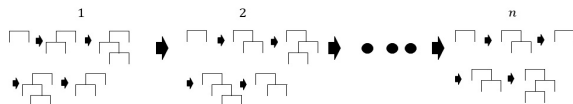


Figure 3: BART - Kindo, Wang e Peña (2016)



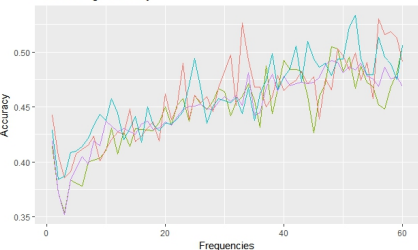
Figure 4: Random Forests - Candell et al. (2016)

# Perguntas

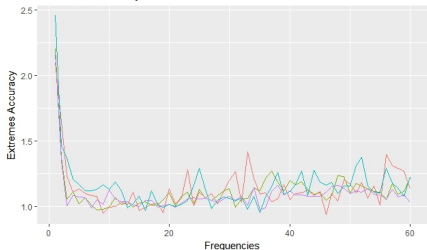
1. Como a agregação temporal afeta a performance dos modelos? (PRADO, 2018)
2. Modificar a definição de retorno nulo de  $R_t = 0$  para  $|R_t| < m$  pode gerar melhorias na previsibilidade dos retornos? (PRADO, 2018)
3. A performance dos modelos é similar e pode gerar resultados financeiros positivos?

# Primeira Pergunta - Análise de Frequência

### Forecasting Accuracy



### Extremes Accuracy

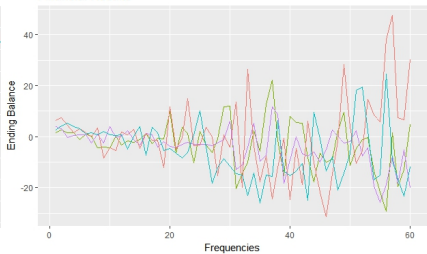


colour  
— BART  
— Boosting Trees  
— Logistic Regression  
— Random Forests

### PETR4.SA



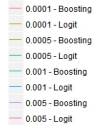
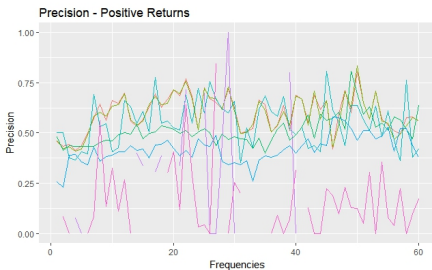
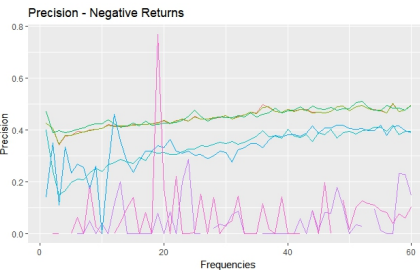
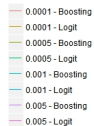
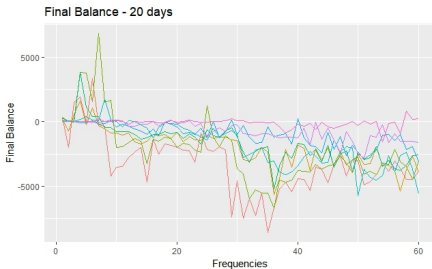
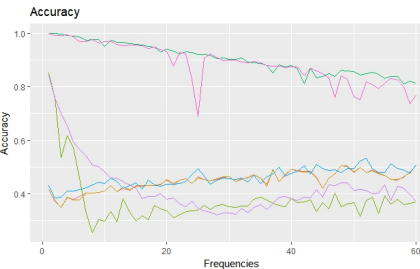
### Financial Results



colour  
— BART  
— Boosting Trees  
— Logistic Regression  
— Random Forests



# Segunda Pergunta - Retorno Nulo



## Primeiras Conclusões

- ▶ Apesar da acurácia global dos modelos aumentar conjuntamente com a agregação temporal, a acurácia nas classes de maior interesse não aumenta junto;
- ▶ Apesar de modelos com grandes bandas intermediárias serem mais acurados novamente a acurácia nas classes de interesse costuma ser maior nas bandas menores.

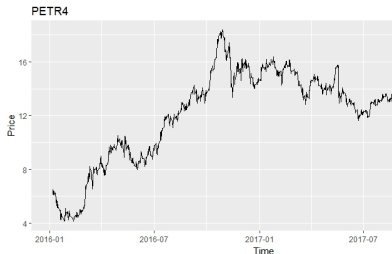
# Terceira Pergunta - Qualidade Preditiva

Table 1: Forecasting Accuracy - 500 days

Algorithm	Boosting Trees	Random Forests	B.A.R.T.	Logit
Accuracy	0.4113	0.4135	0.4124	0.3988
C.I. Superior	0.4142	0.4164	0.4152	0.4016
C.I. Inferior	0.4085	0.4107	0.4095	0.396

Table 2: Forecasting Accuracy - 500 days

Algorithm	Boosting Trees	Random Forests	B.A.R.T.	Logit
Avg. Fin. Ret.	0.00084643	0.00080354	0.00056521	0.00025881
Std. Fin. Ret.	0.0314323	0.0301379	0.0212663	0.00971084



## Principais resultados


- ▶ A característica de que a qualidade dos modelos aumenta conforme se aumenta a frequência não ocorre quando nós olhamos apenas para as classes de interesse;
- ▶ A precisão das classes extremas não aumenta quando nós inserimos um intervalo maior para classificar um retornos como nulo;
- ▶ Todos os modelos de árvores aprendem padrões e se tornam bons preditores com as variáveis de microestrutura oferecidas mas seus retornos médios não compensam os típicos custos de transação.

# Trabalhos Futuros


- ▶ Inclusão de uma estratégia de gestão de risco;
- ▶ Análise do tamanho dos movimentos identificados e não identificados;
- ▶ Análise de Importância das variáveis.

 CANDEL, A. et al. Deep learning with h2o. *H2O. ai Inc*,


2016. 6

 CAO, C.; HANSCH, O.; WANG, X. The information content of an open limit-order book. *Journal of futures markets*, Wiley Online Library, v. 29, n. 1, p. 16–41, 2009.


2

 CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. p. 785–794, 2016.


6

 CHRISTOFFERSEN, P. et al. Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. 2006.

2

 CHRISTOFFERSEN, P. F.; DIEBOLD, F. X. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, INFORMS, v. 52, n. 8, p. 1273–1287, 2006.


2

 DIXON, M. Sequence classification of the limit order book using recurrent neural networks. *Journal of Computational Science*, Elsevier, 2017.


4

 FLETCHER, T.; SHAWE-TAYLOR, J. Multiple kernel learning with fisher kernels for high frequency currency prediction. *Computational Economics*, Springer, v. 42, n. 2, p. 217–240, 2013.


4

 HAN, J. et al. Machine learning techniques for price change forecast using the limit order book data. *Machine learning*, 2015.

4

 KERCHEVAL, A. N.; ZHANG, Y. Modelling high-frequency limit order book dynamics with support vector machines. *Quantitative Finance*, Taylor & Francis, v. 15, n. 8, p. 1315–1329, 2015.

4

 KINDO, B. P.; WANG, H.; PEÑA, E. A. Multinomial probit bayesian additive regression trees. *Stat*, Wiley Online Library, v. 5, n. 1, p. 119–131, 2016.


6

 PRADO, M. L. de. *Advances in financial machine learning*. [S.l.]: John Wiley & Sons, 2018.

7

 TONG, X. T. *Statistical Learning of Some Complex Systems: From Dynamic Systems to Market Microstructure*. Tese (Doutorado), 2013.

3

 TSANTEKIDIS, A. et al. Forecasting stock prices from the limit order book using convolutional neural networks. In: *IEEE Business Informatics (CBI), 2017 IEEE 19th Conference on*. [S.l.], 2017. v. 1, p. 7–12.

4



Federal University of Santa Catarina  
Department of Economics and International Relations

**Rafael Silva Wagner**

M.Sc. Student in Economics  
rafaelsilvawagner@gmail.com

**André Alves Portela Santos**

Ph.D in Quantitative Finance  
andre.portela@ufsc.br