



Aplicação do Support Vector Machine (SVM) em conjunto com variáveis de análise técnica.

Eng. Edson Vinicius Pontes Bastos

UFRJ - Universidade Federal do Rio de Janeiro

engenheiroedsonbastos@gmail.com



Instituto Alberto Luiz Coimbra de
Pós-Graduação e Pesquisa de Engenharia

COPPE
UFRJ



A **força motriz** desta pesquisa surge da necessidade de se investigar a utilidade do *Support Vector Machine* (SVM) para obtenção de retornos significativos.

É notório que os gestores financeiros buscam a todo momento maximizar o retorno de seus investimentos.



Uma tarefa não trivial uma vez que o mercado é dinâmico.

Além disso, vários são os fatores endógenos que interferem nos retornos, por exemplo, fatores políticos, micro e macroeconômicos.



Além disso, em um mercado dito eficiente, retornos anormais são raramente alcançáveis.

Objetiva-se antecipar o movimento, entretanto, em sua maior parte, as séries financeiras apresentam características randômicas.



Portanto, a ideia de que o mercado não pode ser vencido motiva uma controvérsia entre acadêmicos e profissionais de mercado.

Desse modo, apresenta-se a seguinte questão de pesquisa:

Será que a modelagem feita a partir do *Support Vector Machine* (SVM) permite aos investidores obter retornos significativos?



Relacionado ao problema de pesquisa tem-se o seguinte objetivo geral:

Analisar o desempenho dos retornos obtidos por meio do *Support Vector Machine* (SVM) em conjunto com indicadores de análise técnica.



Delineamento dos objetivos específicos:

1. Calcular as variáveis de análise técnica que servirão de entrada para o *Support Vector Machine* (SVM);
2. Delinear uma estratégia de retorno a partir de um modelo de *Support Vector Machine* (SVM);
3. Obter uma modelagem econométrica de retorno por meio de uma regressão linear múltipla;
4. Uso da função *predict* para Predição dos resultados na base de teste
5. Comparar o desempenho das estratégias implementadas: *Support Vector Machine* (SVM) e RL - regressão linear.



A relevância desta pesquisa advém da possibilidade de encontrar uma ferramenta capaz de sinalizar o movimento futuro da ação.

Desse modo será investigada a utilidade de uma estratégia de investimento feita com base em indicadores de análise técnica e *Support Vector Machine*.



Ressalta-se a contribuição desta pesquisa para os investidores individuais.

Sabe-se que o mercado acionário brasileiro é pequeno, quando comparado ao americano, por exemplo, com poucas empresas cobertas por analistas.

Assim, os achados relacionando ferramentas de análise e o desempenho das ações é especialmente importante para investidores sem acesso a analistas de mercado e demais indivíduos que anseiam aprender sobre estratégias para alocação de ativos.



Referencial teórico

Para Lorena e Carvalho (2007), os resultados obtidos por meio do *Support Vector Machine* (SVM) são, por muitas, vezes superiores aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNAs).



Segundo Guenther & Schonlau (2016), analogamente à regressão logística, as SVMs foram inicialmente concebidas para classificações em duas classes.

O problema de classificação envolve as etapas de treinamento e teste, tendo cada conjunto de treinamento um valor de resposta desejado (target), sendo a SVM responsável por fazer a separação através de um hiperplano das distintas classes de dados envolvidas.



1. Coleta da base de dados

2. Cálculo dos indicadores técnicos

3. Obtenção do modelo de SVM.

4. Obtenção do modelo de regressão linear

5. Cálculo dos retornos por SVM, Regressão linear e Buy&hold

6. Análise comparativa do retorno.

Para a coleta de dados utilizou-se o *software R* com o pacote *GETHFDATA* elaborado por Perlin & Ramos (2016), foi considerado o preço de fechamento dentro da janela de 15 minutos, com isso tem-se os indicadores da análise técnica aplicado em candles de 15 min.

O período de análise do preço do ativo ocorreu ao longo do mês de dezembro de 2018. A ação escolhida foi LAME4 entretanto destaca-se que a metodologia se aplica a qualquer outro ativo que tenha liquidez no mercado acionário.



Variáveis do modelo:

Relative Strength Index (RSI) é o **Índice de força relativa (IFR)**, inicialmente proposta por Wilder, (1978) como sendo um indicador técnico do tipo oscilador que afere a relação entre as forças compradoras e vendedoras, este índice varia de 0 a 100.

- Se $RSI(t-1) < 30$ e $RSI(t) > 30$, então compra-se a ação.
- Se $RSI(t-1) > 70$ e $RSI(t) < 70$, então vende-se a ação.



Já o indicador Commodity Channel Index (CCI) é um oscilador de momentum e **mede a variação do preço comparativamente a sua respectiva média.**

Pode-se entender que o mercado está sobre comprado quando está acima de +100, e sobre vendido quando está abaixo de -100.



Segundo Emidio (2016) o objetivo do modelo do SVM é encontrar um hiperplano ótimo, onde possa classificar corretamente todos os pontos referente à etapa de treinamento.

A base de dados foi dividida em duas subamostras: uma de treinamento e outra de teste.

Com o objetivo de obter os parâmetros dos modelos na 1ª subamostra e depois rodar como teste na 2ª subamostra.



Além do SVM também foi utilizado um modelo de regressão linear múltipla para estimar os parâmetros.

Retorno(alvo) = RSL + RsiHigh + RsiLow + CCIH + CCIL

- RsiHigh valores acima de 70; RsiLow valores de RSI abaixo de 30
- CCIHigh valores acima de 120; CCILow valores abaixo de 120

Analogamente estima-se svm da seguinte forma:

SVM Retorno(alvo) = RSL + RsiH + RsiL + CCIH + CCIL,
utilizando-se o kernel linear.



linear: Tabela 01: Parâmetros da regressão

<i>Dependent variable:</i>	
RETORNO (Alvo)	
RSL	129.100*** (1.616)
RsiH	-17.280 (40.221)
RsiL	-131.158** (52.509)
CCIH	-30.349 (58.816)
CCIL	7.830 (41.476)
Constant	9.244 (12.048)

R ²	0.974
Adjusted R ²	0.974
Residual Std. Error	157.756 (df = 216)
F Statistic	1,625.863*** (df = 5; 216)

*p<0.1; **p<0.05; ***p<0.01

Fonte: elaborado pelos autores

Parameters:
 SVM-Type: eps-regression
 SVM-Kernel: linear
 cost: 1
 gamma: 0.2
 epsilon: 0.1

Number of Support Vectors: 16



Observa-se o modelo é estatisticamente significativo como um todo visto que pvalor do teste F é inferior a 5% e possui uma capacidade explicativa de 97,4%. Além disso as variáveis RSL e RsiL são estatisticamente significativas.

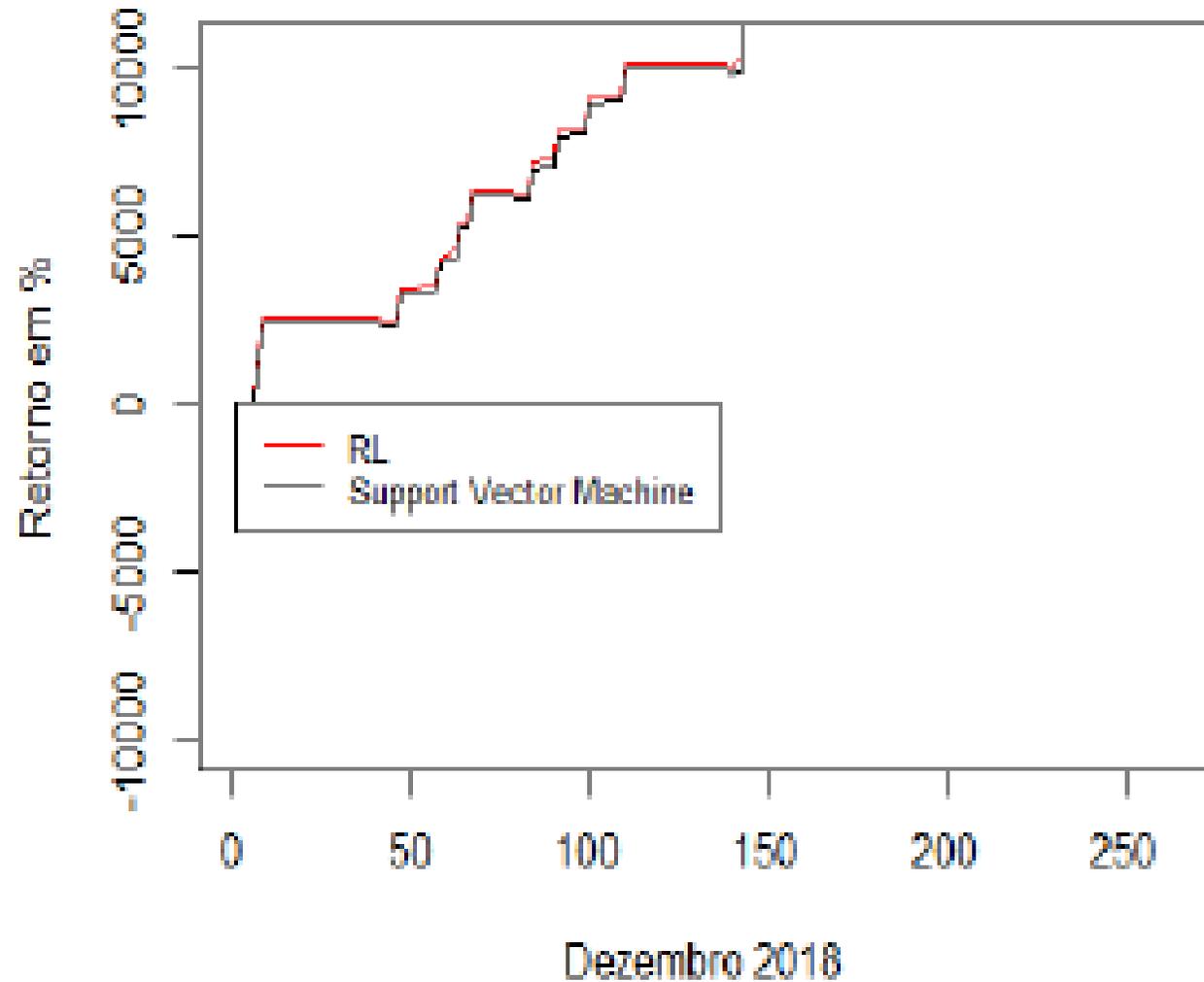
Sobre o teste de multicolineariedade não há evidências estatísticas de que as variáveis preditoras são altamente correlacionadas uma vez que nenhum valor de FIV ficou acima de 5.

Já o teste de homocedasticidade dos resíduos, foram verificados por meio da estatística teste de Breusch-Pagan para a heteroscedasticidade.



Ao final da aplicação do SVM, utilizou-se a função *predict* para Predição dos resultados na base de teste seus respectivos resultados foram confrontados com os resultados obtidos por RL, objetivando verificar o seu desempenho frente conforme exposto no Gráfico 01:

LAME4 - Regressao Linear x Support Vector Machine



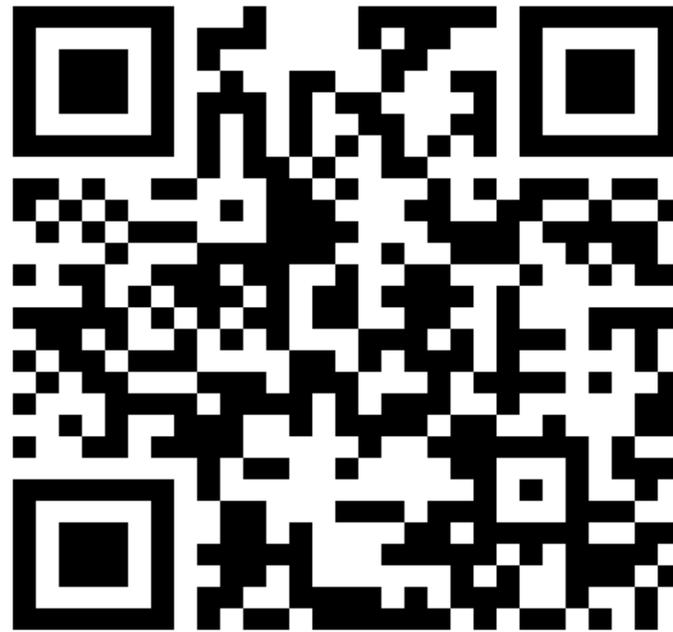


Referencial teórico

- Ebermam, E. (2018). *Desenvolvimento de um Método Híbrido para Negociações de Ações na Bolsa de Valores Brasileira*. 80.
- Emidio, A. (2016). *Booststrap e modelos de support vector machine SVM*. UFPR.
- Emir, S., Dincer, H., & Timor, M. (2012). *A Stock Selection Model Based on Fundamental and Technical Analysis Variables by Using Artificial Neural Networks and Support Vector Machines*. Retrieved from <http://earsiv.beykent.edu.tr:8080/xmlui/handle/123456789/242>
- Fan, A., & Palaniswami, M. (2001). Stock selection using support vector machines. *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222)*, 3, 1793–1798 vol.3. <https://doi.org/10.1109/IJCNN.2001.938434>
- Fávero, L. P. L. (2013). Dados em painel em contabilidade e finanças: teoria e aplicação. *BBR - Brazilian Business Review*, 10(1), 131–156. Retrieved from <http://www.redalyc.org/resumen.oa?id=123025719006>
- Guenther, N., & Schonlau, M. (2016). Support vector machines. *Stata Journal*, 16(4), 917–937. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=120828766&lang=pt-br&site=ehost-live>
- Gujarati, D. N. (2009). *Basic econometrics*. Tata McGraw-Hill Education.
- Lambert, D. R. (1980). Stocks & Commodities. *Commodity Channel Index: Tool for Trading Cyclic Trends*, 1, 120.
- Lorena, A. C., & Carvalho, A. C. P. L. F. de. (2007). Uma Introdução às Support Vector Machines. *Revista de Informática Teórica e Aplicada*, 14(2), 43–67. <https://doi.org/10.22456/2175-2745.5690>
- Perlin, M. S., & Ramos, H. P. (2016). GetHFData: A R package for downloading and aggregating high frequency trading data from Bovespa. *Brazilian Review of Finance*, 14(3), 443–478. Retrieved from <http://bibliotecadigital.fgv.br/ojs/index.php/rbfin/article/view/64587>
- Ritzmann, N. (2016). *Método para Otimização de Janelas de Tempo e Discretização para Classificação de Movimentos Futuros de Ações da Bolsa de Valores* (PhD Thesis). Pontifícia Universidade Católica do Paraná.
- Rodriguez-Lujan, I., Huerta, R., Elkan, C., & Cruz, C. S. (2010). Quadratic Programming Feature Selection. *Journal of Machine Learning Research*, 11(Apr), 1491–1516. Retrieved from <http://www.jmlr.org/papers/v11/rodriguez-lujan10a.html>
- Veropoulos, K., Campbell, C., & Cristianini, N. (1999). *Controlling the Sensitivity of Support Vector Machines*. 6.
- Wilder, J. W. (1978). *New concepts in technical trading systems*.



engenheiroedsonbastos@gmail.com



Bastos

<https://orcid.org/0000-0002-6948-6390>

Obrigado.