

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

Edson Vinicius Pontes Bastos  
UFRJ - Universidade Federal do Rio de Janeiro

Roberto Ivo  
UFRJ - Universidade Federal do Rio de Janeiro

Lino Guimarães Marujo  
UFRJ - Universidade Federal do Rio de Janeiro

### **Resumo:**

Deseja-se investigar a utilidade do algoritmo Support Vector Machine (SVM) para obtenção de retornos significativos. A ideia de que o mercado não pode ser vencido motiva uma controvérsia entre acadêmicos e profissionais de mercado, com isso chega-se a questão de pesquisa: Será que a modelagem feita a partir do Support Vector Machine (SVM) permite aos investidores obter retornos anormais? Para responder a tal questionamento será analisado o desempenho de retornos obtidos por meio do Support Vector Machine (SVM) em conjunto com indicadores de análise técnica. Os resultados indicam que o retorno obtido por RL e por SVM é positivo na janela em estudo. Com isso apresentam-se evidências de que o Support Vector Machine (SVM) é eficaz e o seu emprego permite aos investidores obter retornos anormais.

**Palavras Chaves:** Support Vector Machine, high frequency data, market microstructure.

### **1.Introdução**

A força motriz desta pesquisa surge da necessidade de se investigar a utilidade do algoritmo *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, portanto, vencer o mercado é o objetivo. Entretanto, em um mercado dito eficiente, retornos anormais são raramente alcançáveis. 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. Para cumprimento do objetivo geral torna-se necessário o delineamento dos objetivos específicos a saber:

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. Comparar o desempenho das estratégias implementadas: Support Vector Machine (SVM), RL - regressão linear.

Uma das condições para obtenção de retornos anormais é a de que a informação contida na análise técnica antecipe o futuro econômico do ativo em questão, fato este que será precificado pelo mercado gerando, assim, retornos positivos.

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). Destarte os autores ressaltam a importância da metodologia SVM.

Com isso, 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*.

Logo, este estudo é uma forma de avaliar a relevância da metodologia de SVM para o mercado brasileiro que possui características institucionais distintas do mercado americano. Outrossim, pesquisas a respeito de ferramentas para análise de investimento são relevantes para instituições privadas, como: bancos, corretoras e fundos privados de investimento e instituições públicas, como as estatais que se utilizam de estratégias de investimento para alocação e diversificação de suas carteiras dos fundos de pensão, por

exemplo. E por fim, porém não menos importante, ressalta-se a contribuição desta pesquisa para os investidores individuais.

Portanto, a pesquisa destina-se a servir como fonte bibliográfica para investidores privados, analistas, consultores, e pessoas com funções de tomada de decisão para investimento. 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 em carteiras de investimento.

## 2. Referencial teórico

A metodologia *Support Vector Machine* é utilizada em engenharia econômico financeira para formação de carteiras de ações ou ainda para prever movimentos de mercado, e é neste segundo ponto que se aplica *Support Vector Machine* neste trabalho. Segundo Guenther & Schonlau (2016), analogamente à regressão logística, as SVMs foram inicialmente concebidas para classificações em duas classes. Esta abordagem foi posteriormente ampliada para contemplar resultados contínuos e classificações com mais de duas classes.

Inicialmente empregado por Fan e Palaniswami (2001), objetivando retornos anormais, ou seja, acima do índice de mercado, com uso de variáveis contábeis e de preço das ações negociadas na Bolsa de Valores da Austrália, Fan e Palaniswami (2001) empregaram o SVM para identificar ações com potencial de superar o índice de mercado. A carteira igualmente ponderada formada pelas ações selecionadas com uso do SVM apresentou um retorno total de 208% ao longo de um período de cinco anos, e, portanto, superou significativamente o índice de mercado que era 71%.

Fan e Palaniswami (2001) agruparam as variáveis contábeis em oito categorias: retorno sobre o capital, rentabilidade, alavancagem, investimento, crescimento, liquidez de curto prazo, retorno sobre o investimento e risco. Desse modo, para o conjunto de variáveis que compunham cada uma das categorias Fan e

Palaniswami (2001) efetuaram uma Análise de Componentes Principais e definiram o primeiro componente principal como a variável representativa do grupo da categoria financeira, tal indicador foi empregado na formação da carteira de ações.

A variável dependente foi definida pelos autores como o preço dos ativos, os quais foram classificados como superior - Classe 1 se tais papeis estivesse acima do terceiro quartil empírico para o retorno, e inferior - Classe 2 para as piores ações, ou seja, aquelas que possuíram retorno inferior ao primeiro quartil empírico da distribuição de retornos empresas listadas na Bolsa de Valores Australiana. Entretanto, como o número de observações incluídas na Classe 1 e Classe 2 eram distintos, os autores precederam com uma abordagem de SVM para dados não balanceados. Seguindo a metodologia proposta por Veropoulos, Campbell e Cristianini (1999) que tem como princípio a modificação do problema dual clássico de SVM.

Com o objetivo de fornecer uma modelagem melhor para os retornos, Emir, Dincer e Timor (2012) integraram indicadores técnicos e fundamentalistas e compararam o SVM com as RNA em quatro cenários: somente indicadores técnicos; indicadores fundamentalistas; indicadores técnicos e fundamentalistas e seleção específica de indicadores para a análise seguindo o método proposto por Rodriguez-Lujan, Huerta, Elkan, e Cruz (2010), objetivando analisar qual metodologia apresentaria o maior poder explicativo, os autores empregaram Rede Neural Artificial (RNA), uma abordagem amplamente utilizada para problemas de classificação financeira em conjunto com Support Vector Machines (SVM), nesta pesquisa a população em estudo foi composta por todas as ações que compunham o Índice Istanbul Stock Exchange (ISE) para o período de 2002 a 2010.

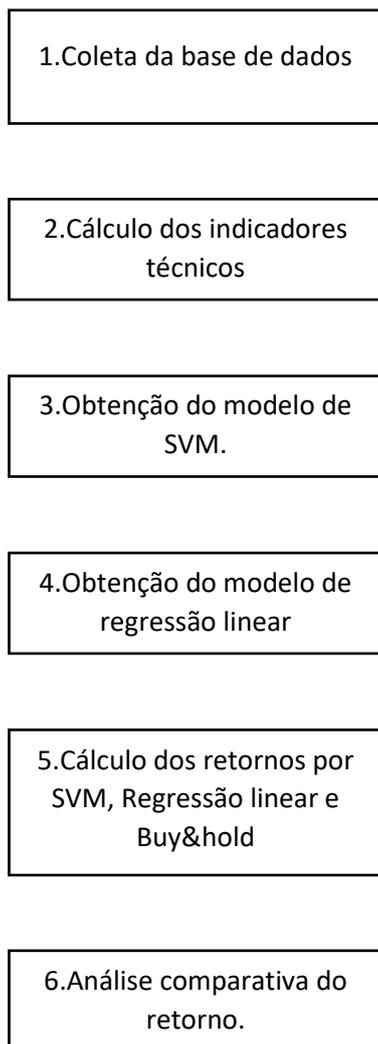
Segundo Emir, Dincer e Timor (2012), o Support Vector Machine apresenta resultados promissores de um novo método de machine learning. Além disso, também foi verificado que os modelos construídos usando o método SVM nos dois tipos de análises mostraram resultados melhores do que RNA, visto que o SVM apresentou superioridade na acurácia de previsão para a maioria das comparações realizadas, além disso, o núcleo que apresentou o

melhor desempenho foi o núcleo Gaussiano para o SVM. Com estes indicativos da potencialidade do SVM em engenharia econômica financeira, motiva-se a investigação mais aprofundada desta metodologia no contexto financeiro brasileiro.

### 3. Delineamento metodológico

A metodologia da pesquisa segue o fluxograma:

Figura 01: Fluxograma



Fonte: Elaborado pelos autores

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 o papel preferencial da empresa Lojas

Americanas do setor de bens de consumo cuja sigla é LAME4. Entretanto destaca-se que a metodologia se aplica a qualquer outro ativo que tenha liquidez no mercado acionário.

A variável 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 de um determinado papel, este índice varia de 0 a 100.

Para Ebermam (2018), tal indicador sinaliza regiões de sobrecompra e sobrevenda, quando o indicador está abaixo de um limite, tem-se o *RSILow*, comumente considera-se este nível igual a 30, neste cenário interpreta-se que o preço da ação está em uma zona de sobrevenda. Ebermam (2018) defende então que a força vendedora está perdendo força e isso pode ser um sinal que o preço da ação irá subir.

De modo análogo, acima do limite superior considerado normalmente 70, tem-se o *RSIHigh*, desta forma o preço está na zona de sobrecompra, indicando que a força compradora está se reduzindo, com isso aumenta-se a probabilidade de que o preço da ação caia.

Resumidamente, o indicador mede a aceleração do movimento das cotações de um determinado ativo, uma vez que é preciso desacelerar primeiro para só então depois modificar a direção. Com isso, tal métrica fornece suas indicações à medida que o movimento diminui sua velocidade. simplificada tem-se o seguinte padrão a ser seguido:

- 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. Conforme definido por Lambert (1980), atribuiu-se uma constante no igual a 0,015 para a multiplicação com o desvio padrão. Tal constante assegura que cerca de 70% a 80% dos valores se situam entre -100 e +100.

Segundo Ritzmann (2016) para a correta compreensão do indicador CCI é

necessário compreender os conceitos de sobre comprado e sobre vendido. Como sobre comprado entende-se por um período de dias subsequentes em que houve uma alta expressiva sem baixas relevantes para amenizar a alta e permitir a entrada de novos compradores. Já sobre vendido é o inverso, ou seja, uma ação sobre vendida sofreu quedas subsequentes sem haver altas significativas.

Uma ação sobre comprada ou sobre vendida pode indicar que o movimento está próximo de se encerrar. Para interpretar os resultados do indicador CCI utiliza-se deste conceito, pode-se entender que o mercado está sobre comprado quando está acima de +100, e sobre vendido quando está abaixo de -100.

Entretanto, alguns investidores utilizam a movimentação destes valores para entender a força do mercado. O rompimento do valor de +100 para cima pode representar força na tendência de alta, porém quando volta abaixo de +100 entende-se que o mercado está corrigindo sua alta recente, o mesmo vale para valores abaixo de -100. (Ritzmann, 2016).

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 e, além disso, com a maior margem de separação possível para os vetores da classe -1 e da classe +1.

Além do SMV também foi utilizado o modelo econométrico de regressão linear múltipla e, seguindo as orientações de Gujarati (2009) e Fávero (2013), para estimar as funções de regressão, utilizou-se o método dos Mínimos Quadrados Ordinários (MQO), onde a probabilidade de erro tipo I, para análise dos pressupostos foi definida em 5%, dessa forma tem-se um índice de confiança (IC) de 95%.

A análise do coeficiente R<sup>2</sup> será validado por meio do p-valor do teste F, já para validação dos coeficientes dos regressores será utilizado o p-valor do teste T, sendo analisado sobre os aspectos de 1% 5% e 10 % de significância. A seguir apresenta-se a equação utilizada para a modelagem econométrica.

$$\text{Retorno(alvo)} = \text{RSL} + \text{RsiHigh} + \text{RsiLow} + \text{CCIH} + \text{CCIL}$$

RsiHigh valores de RSI acima de 70

RsiLow valores de RSI abaixo de 30

CCIHHigh valores acima de 120

CCILow valores abaixo de 120

RSL é a média móvel de 8 períodos

Analogamente estima-se svm da seguinte forma:

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

A base de dados foi dividida em duas subamostras: uma de treinamento e outra de teste, conforme prática habitual de machine learning, com o objetivo de obter os parâmetros dos modelos na 1ª subamostra e depois rodar como teste na 2ª subamostra.

#### 4. Análise dos resultados

Após a coleta de dados seguiu-se com a obtenção dos modelos de regressão, suas respectivas saídas encontram-se dispostas na tabela 1:

Tabela 01: Parâmetros da regressão linear:

<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)
Observations	222

R <sup>2</sup>	0.974
Adjusted R <sup>2</sup>	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

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, tendo o RSL apresentado coeficiente positivo e o RsiL negativo.

Com relação aos testes dos pressupostos para validação do modelo de regressão, foi realizado teste de normalidade dos resíduos, onde na Hipótese nula afirma-se que o erro tem distribuição Normal, ao nível de significância de 5% rejeita-se  $H_0$  na equação 1, porém pelo teorema do limite central, como a amostra é superior a 30, pode-se relaxar com o pressuposto de normalidade e seguir adiante com a análise do modelo.

Sobre o teste de homocedasticidade dos resíduos, foram verificados por meio da estatística teste de Breusch-Pagan para a heteroscedasticidade, onde em sua Hipótese nula afirma-se que o modelo é sem heteroscedasticidade, ou seja, é desejado que a variância dos resíduos seja constante para todos os valores.

Ao nível de significância de 5% rejeita-se a Hipótese nula, portanto, segue-se com a obtenção do modelo por meio da correção de erros padrões robustos à heteroscedasticidade.

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

Tabela 02 : Resultados do modelo SVM:

Parameters:

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

Number of Support Vectors: 16

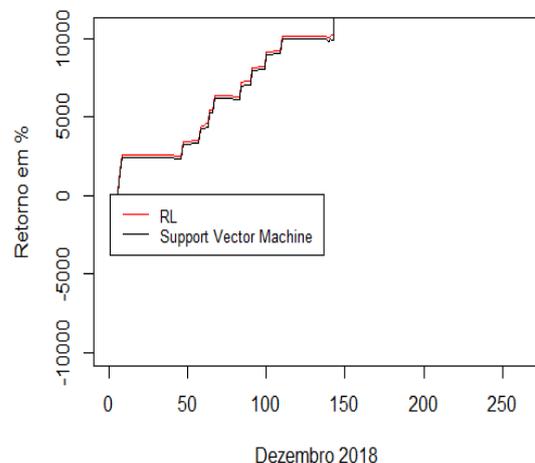
Fonte: elaborado pelos autores

## 5. Conclusão

Ao final da aplicação do SVM, seus respectivos resultados foram confrontados com os resultados obtidos por metodologias já existentes e consolidadas na literatura de finanças, objetivando verificar o seu desempenho e sua validade frente aos demais métodos, este resultado comparativo encontra-se exposto no Gráfico 01:

**Gráfico1: Comparação das metodologias**

**LAME4 - Regressao Linear x Support Vector Machine**



Eixo y : Retorno Obtido

Eixo X: Período em análise (unidades coletadas a cada de 15 minutos).

Fonte: Elaborado pelos autores

Nota-se que o retorno obtido por RL e por SVM é superior à metodologia de investimento Buy&Hold, uma vez que seu retorno foi superior na janela em estudo a saber: dezembro de 2018. Desse modo apresentam-se evidências de que o *Support Vector Machine* (SVM) é eficaz e o seu

emprego permite aos investidores obter retornos significativos.

## 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*. Retrieved from <http://agris.fao.org/agris-search/search.do?recordID=US201300554903>