

## **Tomada de Decisão de Investimento em Ativos da Bolsa de Valores de São Paulo Baseado na Técnica Support Vector Machine**

### **Autoria**

Felipe Dias Paiva - fpavia@dcsa.cefetmg.br

Prog de Pós-Grad em Admin - PPGA/CEFET-MG - Centro Federal de Educação Tecnológica de Minas Gerais

Wendel Moreira Duarte - wmoreiraduarte@gmail.com

Prog de Pós-Grad em Admin - PPGA/CEFET-MG - Centro Federal de Educação Tecnológica de Minas Gerais

Marcília Junia de Andrade Oliveira - marcilianaandrade@hotmail.com

Prog de Pós-Grad em Admin - PPGA/CEFET-MG - Centro Federal de Educação Tecnológica de Minas Gerais

Juliano Lima Pinheiro - jlp@gold.com.br

Prog de Pós-Grad em Admin - PPGA/CEFET-MG - Centro Federal de Educação Tecnológica de Minas Gerais

### **Agradecimentos**

CAPES; CNPq; CEFET-MG

### **Resumo**

O presente trabalho teve como objetivo principal compor um modelo baseado na técnica support vector machine (SVM) para tomada de decisão de investimento. A SVM foi utilizada com o intuito de classificar os ativos com potencial de alcançar um determinado ganho proposto. Para execução da pesquisa, foi definido o modelo em ativos listados no índice Ibovespa. O período explorado compreendeu de 04/01/2010 a 28/12/2016. O resultado da pesquisa apresentou um resultado bastante interessante à medida que o retorno de investimento baseado na SVM foi superior ao baseline (Ibovespa). Aponta-se também que, construir um classificador com ativos homogêneos quanto ao setor econômico de atuação, proporcionou um melhor índice de rentabilidade, menor cardinalidade média das carteiras de investimento, menor máximo dropdown.

## Tomada de Decisão de Investimento em Ativos da Bolsa de Valores de São Paulo Baseado na Técnica *Support Vector Machine*

**RESUMO:** O presente trabalho teve como objetivo principal compor um modelo baseado na técnica *support vector machine* (SVM) para tomada de decisão de investimento. A SVM foi utilizada com o intuito de classificar os ativos com potencial de alcançar um determinado ganho proposto. Para execução da pesquisa, foi definido o modelo em ativos listados no índice Ibovespa. O período explorado compreendeu de 04/01/2010 a 28/12/2016. O resultado da pesquisa apresentou um resultado bastante interessante à medida que o retorno de investimento baseado na SVM foi superior ao *baseline* (Ibovespa). Aponta-se também que, construir um classificador com ativos homogêneos quanto ao setor econômico de atuação, proporcionou um melhor índice de rentabilidade, menor cardinalidade média das carteiras de investimento, menor máximo dropdown.

**Palavras-chave:** Investimento; *support vector machine*; setor econômico.

### 1 INTRODUÇÃO

A previsão dos preços das ações é considerada uma das tarefas mais desafiadoras do processo que lida com séries temporais financeiras, pois o mercado de ações é essencialmente dinâmico, complexo, evolucionário, não linear, nebuloso, não paramétrico e caótico por natureza. Além disso, o mercado de ações é extremamente sensível a fatores políticos e a condições micro e macroeconômicas, bem como a expectativas e inseguranças dos investidores. (Ballings, Van den Poel, Hespels, & Gryp, 2015; Kara, Boyacioglu, & Baykan, 2011; Tan, Quek, & Ng, 2007).

Para a *mainstream* da teoria financeira, a previsão de preços de ativos financeiros é impossível. De acordo com o principal pilar teórico dessa corrente – a Hipótese de Mercado Eficiente (HME) –, a tarefa de prever preços futuros tendo como base o comportamento passado de um ativo financeiro não passa de um exercício ineficaz para o alcance de retornos anormais, dado que a função de distribuição de uma série financeira denota um movimento *browniano*, que é concebido por características randômicas, independentes e de distribuição *gaussiana*.

Em contrapartida, existem estudos que rechaçam a HME, sustentando que o mercado acionário na verdade não se estabelece como aleatório e que as séries financeiras possuem memória de longo prazo. Desta forma, uma vez identificado o seu padrão de comportamento, torna-se factível delinear algum modelo de previsibilidade (C. Huang, Yang, & Chuang, 2008; Lo, Mamaysky, & Wang, 2000; Malkiel, 2003; Mandelbrot & Hudson, 2004; Patel, Shah, Thakkar, & Kotecha, 2015). Já com relação às características gerais das séries, parece existir uma concordância quanto à complexidade e nebulosidade das mesmas, sendo indicado, portanto utilizar técnicas robustas e apropriadas para manuseá-las (Brabazon & O’Neill, 2006).

Assim, sistemas especialistas têm a sua utilização cada vez mais demandada, como resposta às características do ambiente. Persiste, assim, o desafio de identificar um nexo entre passado e futuro, com o objetivo de prever o preço ou o retorno de uma ação. Para esse efeito, destacam-se dois campos investigativos (Sheta, Ahmed, & Faris, 2015):

I) Modelos econométricos: tratam-se de modelos estatísticos tais como regressão linear, auto-regressivo (AR), auto-regressivo de média móvel (ARMA), auto-regressivo com heterocedasticidade (ARCH), auto-regressivo com heterocedasticidade generalizado (GARCH) etc. Um ponto-chave de avaliação do uso desses modelos centra-se nos pressupostos que devem ser atendidos pelas séries financeiras a fim de garantirem a qualidade e confiabilidade dos resultados produzidos.

II) Modelos baseados em *soft computing*: tratam-se de modelos baseados em técnicas de inteligência artificial, e apresentam-se como bastante interessantes graças à capacidade de tratarem dados complexos, imprecisos e volumosos. Essas características de dados, quando aplicadas em outros modelos, tendem a obscurecer o significado subjacente, bem como restringir a obtenção de informações úteis. Além disso, há técnicas de inteligência artificial que permitem utilizar dados de diferentes naturezas (qualitativos e quantitativos). Sem contar que não estão sujeitos à rigidez de pressupostos, como os que são impostos aos modelos econométricos.

Diante desses pontos, parece-nos adequada a utilização de um modelo ancorado em alguma técnica de inteligência computacional. Assim, ao propormos o desenvolvimento de um modelo de tomada de decisão baseado no uso de técnicas de sistemas inteligentes, buscamos explorar o grande potencial que essas técnicas têm para lidar com problemas relacionados à área financeira, dada a capacidade destas de gerir sistemas com arquiteturas complexas de dados. De acordo com Mandelbrot e Hudson (2004), a finança “pós-moderna” baseada na inteligência do silício, tem ainda um longo caminho a percorrer na busca de soluções convincentes, sendo que as finanças computacionais integram explicitamente matemática, estatística, finanças, economia, sistema de informação e computação.

Na execução desta pesquisa optamos por utilizar a técnica de inteligência computacional, a *Support Vector Machine* (SVM). A SVM tem tido seus resultados enaltecidos graças à sua capacidade de controle da função de decisão e resistente ao problema de *overfitting*, alcançando, assim, um alto desempenho de generalização (Burgess, 1998).

Desta forma, o presente trabalho teve como objetivo principal estruturar um modelo classificador baseado em SVM para prever o movimento de alta dos ativos do IBovespa. Como objetivos específicos, definiu-se por: (I) simular os efeitos quanto a rentabilidade, a cardinalidade das carteiras e exposição ao risco que investidor teria caso optasse por investir com base no classificador desenvolvido; (II) avaliar os efeitos do modelo para um conjunto de ativos agrupados de acordo com o setor econômico da empresa que ele representa.

## **2 SUPPORT VECTOR MACHINE**

A técnica SVM surgiu na metade dos anos de 1990, fruto do emprego prático dos conceitos desenvolvidos pela Teoria de Aprendizado Estatístico. O pesquisador de maior destaque para desenvolvimento da SVM é o russo Vladimir N. Vapnik, que esteve à frente da discussão desde o final da década de 1960, quando surgiram os primeiros estudos da Teoria de Aprendizado Estatístico.

A SVM objetiva resolver problemas de ordem de reconhecimento de padrões, classificação, estimativa de regressão e séries temporais, além de estimação de densidade (W. Huang, Nakamoria, & Wang, 2005; Vapnik, 1999). Diversos estudos têm destacado as potencialidades de utilização das ferramentas preditivas da SVM em questões relacionadas ao mercado de ações, como mostrado na Tabela 1.

De acordo com Vapnik (1999), a ideia básica da SVM é mapear os vetores de entrada em um espaço de grande dimensão por meio de algum elemento de mapeamento não linear a ser definido. Nesse espaço, um hiperplano ótimo de separação é construído (Figura 1).

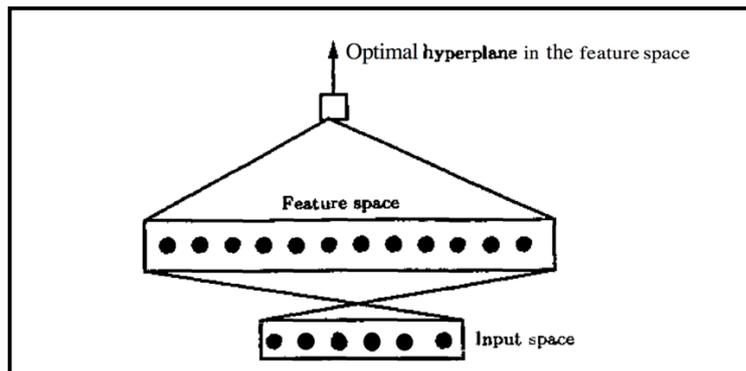


Figura 1- Hiperplano ótimo

A ideia da SVM é aplicar conceitos de um modelo linear para separar os *inputs* a partir de um mapeamento não linear em um espaço característico de alta dimensão. O modelo linear construído no novo espaço pode agora representar um limite de decisão não linear frente ao espaço original. A SVM focará, assim, em estabelecer hiperplanos de separação ótima, sendo que os pontos de treinamento ( $x_i \rightarrow y_i$ ) que estão mais próximos do hiperplano de separação ótimo são denominados de vetores de suporte e estabelecem o limite do plano de decisão. Em casos gerais em que os dados não são linearmente separados, a SVM utiliza máquinas não lineares para encontrar um hiperplano que minimiza o número de erros no conjunto de treinamento (Ding, Song, & Zen, 2008).

Assim, vamos supor um conjunto de pontos para treinamento,  $D = \{x_i, y_i\}_{i=1}^N$ , onde o vetor de *inputs*  $x_i = (x_i^{(1)}, \dots, x_i^{(n)}) \in R^n$  e os outputs  $y_i \in \{0, 1\}$ , são satisfeitos pela seguinte condição:

$$y_i(w^T \phi(x_i) + b \geq 1) \quad i = 1, \dots, N$$

Onde  $w$  representa o vetor peso e  $b$  a variável *bias*. A função não linear  $\phi(\cdot): R^n \rightarrow R^{nk}$  mapeia os *inputs* no espaço em alta dimensão. Assim, analisando a equação acima, pode-se perceber que há várias soluções para separar os pontos de treinamento. A Figura 2 corrobora o entendimento.

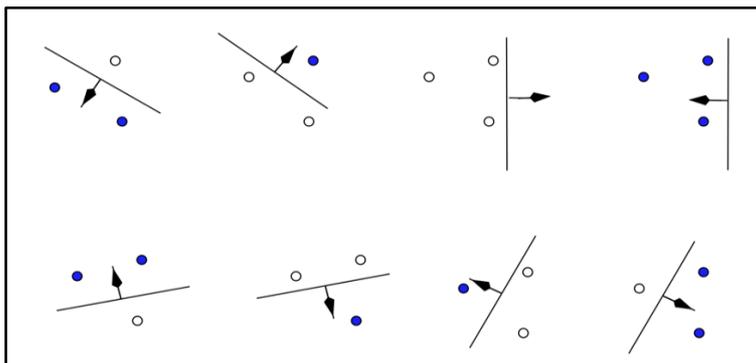


Figura 2 - Separação de conjuntos de pontos por hiperplano

De uma perspectiva de generalização, é melhor escolhermos dois hiperplanos limitadores em lados opostos de um hiperplano de separação  $w^T \phi(x) + b = 0$  com maior margem  $2/(\|w\|^2)$ .

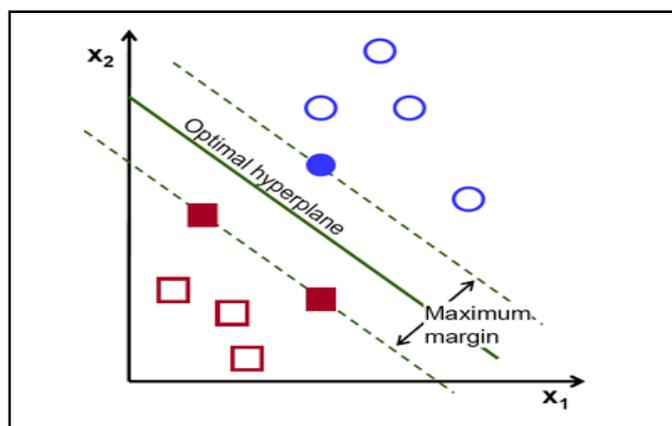


Figura 3 - Optimal separating hyperplane

Contudo, vários problemas de classificação são casos linearmente não separáveis. Portanto, é necessário introduzirmos variáveis de folga ( $\xi_i$ ) para permitirmos a classificação errada. Assim, o problema de otimização torna-se:

$$\text{Min}_{w, b, \xi} \left( \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \right)$$

$$\text{Sujeito: } \begin{cases} y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i & i = 1, \dots, N \\ \xi_i \geq 0 \end{cases}$$

Onde  $C$  é o parâmetro de penalidade do termo de erro. A solução do problema primário é obtida após a construção Lagrangiana. Em seguida, o problema primal pode ser convertido no seguinte QP-problema:

$$\text{Max}_{\alpha} \left( \alpha^T \alpha - \frac{1}{2} \alpha^T Q \alpha \right)$$

$$\text{Sujeito: } \begin{cases} 0 \leq \alpha_i \leq C, i = 1, \dots, N \\ \sum_{i=1}^N \alpha_i y_i = 0 \end{cases}$$

Onde  $\alpha_i$  é um multiplicador Lagrangiano,  $Q_{ij} = y_i y_j \phi(x_i)^T \phi(x_j)$ . Em razão do esforço computacional exigido, o produto interno é substituído pela função kernel, que satisfaz a condição de Mercer,  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ . Finalmente, alcançamos a função de decisão não linear no espaço primal para o caso linearmente não separável:

$$y(x) = \text{sgn} \left( \sum_{i=1}^N \alpha_i y_i K(x_i, x_j) + b \right)$$

As funções do kernel que são usadas no SVM são:

- Kernel linear:  $K(x_i, x_j) = x_i^T x_j$ ;
- Kernel polinomial:  $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$ ;
- Kernel radial:  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ ;
- Kernel sigmoid:  $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$ .

Algumas características da SVM merecem destaque:

- A convexidade da função objetivo como uma das vantagens da SVM, dado que o treinamento da SVM equivale à solução de um problema de programação quadrática, no qual a solução do problema será sempre única e globalmente melhor. Já a RNA demanda uma otimização não linear com o risco, sempre eminente, de o algoritmo ficar refém de mínimos locais. (Ballings et al., 2015).
- A SVM possui um maior poder de precisão se comparada a outros métodos individuais de previsão. Destaca-se que a SVM atinge uma melhor performance em decorrência de ter sido concebido para minimizar o risco estrutural, ao passo que as outras técnicas focam a minimização do risco empírico. Ou seja, a técnica SVM busca minimizar o limite superior do erro de generalização em detrimento ao erro de treino (W. Huang et al., 2005; Sands, Tayal, Morris, & Monteiro, 2015).
- A SVM é capaz de processar grandes volumes de dados de forma robusta, sem que haja ocorrência de *overfitting* (Awad & Khanna, 2015).

A Tabela 1 apresenta um resumo de vários trabalhos relevantes encontrados na literatura que utilizaram a SVM em previsões de séries financeiras de diferentes mercados e para períodos distintos. Todos esses trabalhos destacam o grande potencial de aplicação da SVM para problemas dessa natureza.

**Tabela 1- Trabalhos relacionados**

Referência	Amostra	Resultados
Pan, Xiao, Wang e Yang (2017)	A amostra foi composta por dados do índice S&P 500 de junho/2008 a abril/2015.	O trabalho usa um modelo SVM usado na previsão dos movimentos de preço que considera informações em diversas frequências. Os resultados indicam que o modelo supera as referências propostas, se mostrando uma opção interessante para aplicações práticas.
Li et al. (2016)	Amostra composta por preço e notícias de 23 ações do Hang Seng Index no ano de 2001.	Foram feitas previsões usando as informações de preços e notícias para o mercado em questão. Os resultados usando SVM e ELM superaram a BP-NN em termos de acurácia e velocidade de processamento.
Thenmozhi e Sarath Chand (2016)	Amostra composta de seis mercados globais (Dow Jones, S&P 500, FTSE, NSI, SGX, Hang Seng e Shanghai Stock Exchange) entre 1999 e 2011.	As previsões obtidas com os modelos usando SVM superaram previsões usando modelos de regressão e indicadores de análise técnica em todos os sete mercados analisados para todo o período analisado.
Patel et al. (2015)	Foram usados dados de duas ações e dois índices do mercado indiano (CNX e BSE) de 2013 a 2012.	Foram comparadas as técnicas de random forest, naive-Bayes, ANN e SVM. Para dados contínuos, a técnica de random forest obteve melhor resultado, para dados discretos, a técnica naive-Bayes alcançou maior acurácia.
Zbikowski (2015)	Amostra composta por 20 ações do índice S&P 500 escolhidas aleatoriamente de janeiro/2003 a outubro/2013.	O trabalho usa um modelo SVM usado na previsão de tendências de curto prazo no mercado financeiro. Os resultados mostram que o modelo apresenta melhor retorno e drawdown durante o período de análise.
Fenghua, Jihong, Zhifang e Xu (2014)	A amostra foi composta ativos listados no Shanghai Stock Exchange de janeiro/2009 a setembro/2013.	Foram usadas técnicas de singular spectrum analysis (SSA) e ensemble empirical mode decomposition (EEMD) para análise prévia das informações. Posteriormente, os dados foram acoplados ao modelo SVM e foram feitas as comparações. Concluiu-se que o modelo SSA-SVM obteve melhor acurácia que os demais modelos.
Yu Chen e Zhang (2014)	A amostra considera dados do Shanghai Stock Exchange Index de 2009 a 2010.	Foram usadas técnicas de análise componentes principais (PCA) para análise prévia das informações. Em seguida, os dados foram acoplados ao modelo SVM e foram feitas as comparações. Foi alcançada um acurácia de 61.79% na análise out-of-sample.
Luo e Chen (2013)	Amostra composta por 20 ações do Shanghai Stock Exchange Index para o período de janeiro/2010 a agosto/2011.	Foram comparados modelos de piecewise linear representation (PLR) acoplados com back-propagation artificial neural network (BPN) e weighted support vector machine (WSVM). Os resultados mostram que o modelo LPR-WSVM é superior em termos de acurácia e rentabilidade em todos os períodos.
Zhiqiang, Huaiqing e Quan (2013)	A amostra considera dados do Shanghai Stock Exchange Index e Dow Jones Index de 2000 a 2004.	Foram usadas técnicas de locality-preserving projection (LPP) e particle swarm optimization (PSO) para análise prévia das informações. Em seguida, os dados foram acoplados ao modelo SVM e foram feitas as comparações. Foi alcançada um acurácia de 61.73% para SSE e 57.94% para DJI, na análise out-of-sample.
Kara et al. (2011)	A amostra foi composta por dados de preço do ISE National 100 Index de 1997 a 2007.	O trabalho apresenta dois modelos para previsão do movimento de preços com base em indicadores de análise técnica, um usando artificial neural networks (ANN) e outro usando support vector machines (SVM). Os resultados indicam superioridade em termos de acurácia do modelo ANN (75.74%) em relação ao SVM (71.52%).
Ni, Ni e Gao (2011)	Amostra composta por preços do Shanghai Stock Exchange Index de 2000 a 2008.	O trabalho usa fractal feature selection method para escolha de features do modelo SVM usado na previsão diária de preço. Os resultados mostram que esse processo de escolha seleciona um número um pouco que menor que os demais métodos de escolha com melhores valores de acurácia.
Wen, Yang, Song e Jia (2010)	Foram selecionadas 442 ações do S&P 500 de março/2004 a outubro/2005.	O trabalho propõe um modelo inteligente de negociação de ações baseado em oscilações do mercado, combinando stock box theory and SVM. Os resultados se mostraram que o modelo supera significativamente a estratégia buy-and-hold.
Lee (2009)	O objetivo do trabalho foi de prever os movimentos de preços no NASDAQ index. Para tal foram selecionados dados do período de novembro/2001 a novembro/2007.	O trabalho usa um método híbrido para escolha de features do modelo SVM usado na previsão dos movimentos de preço. Os resultados mostram que o modelo tem melhores resultados que modelo SVM com outros métodos para escolha de features e que modelo de rede neural artificial com retro propagação.
Huang et al. (2005)	A amostra foi composta por dados do índice NIKKEI 225 de janeiro/1990 a dezembro/2002.	O trabalho usa um modelo SVM usado na previsão dos movimentos de preço. Os resultados mostram que o modelo SVM supera outros métodos de classificação.

### 3 METODOLOGIA

Para desenvolver a presente pesquisa utilizamos ativos listados na BMF Bovespa. Sendo que, para uma delimitação melhor da amostra, foi definido a utilização das ações participantes do índice da Bolsa de Valores de São Paulo (IBOV) em 31/01/17, que tiveram negociações diárias desde o ano de 2010. Estabelecidos os critérios de amostragem, pode-se atestar por meio do Tabela 2, no Apêndice, as 50 empresas que compuseram a amostra.

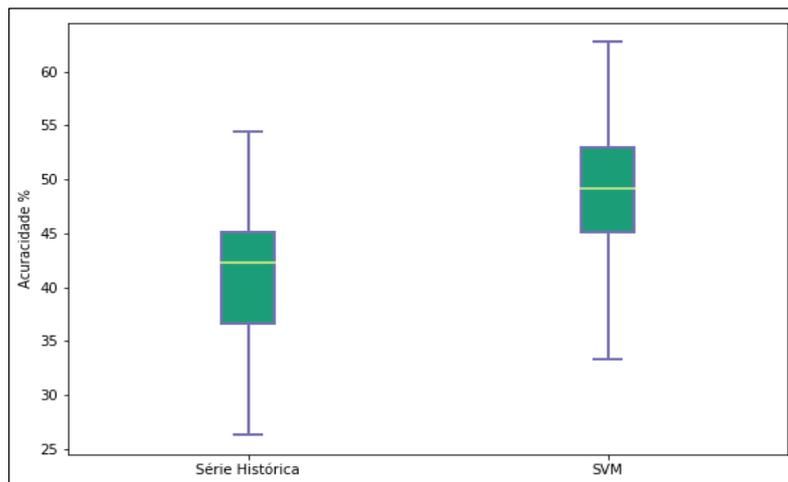
Uma vez definidas as empresas, recorreu-se ao terminal da Bloomberg para coletar as cotações diárias ajustadas de preços de abertura, fechamento, máximo, mínimo e volume, de 04/01/2010 a 29/12/2017. Essas informações serviram para que se preparasse os *inputs* do modelo de classificação baseado em SVM. Trabalhamos com o sistema de janelas deslizantes diárias para previsão do movimento dos ativos. Uma vez classificadas as ações entre as com potencial de atingir o lucro desejado e aquelas não tinham, utilizamos essa informação para simular a compra de ativos. Ficou estabelecido que a proporção de investimento em cada sinal de ativo seria 1/N. Para avaliar se alcançaríamos melhores decisões de investimento operando com ativos agrupados a partir dos seus respectivos setores econômicos, recorreu-se a classificação da Global Industry Classification Standard (GICS) acessado via Bloomberg. A pesquisa foi realizada utilizando o software *Python 3.6*.

### 4 RESULTADOS

Uma vez tratados os dados base para a realização da pesquisa, iniciou-se o desenvolvimento treinando o algoritmo de classificação baseado na técnica SVM para previsão do movimento dos ativos. Realizada a simulação não supervisionada de classificação, comparou-se a performance da SVM com os dados reais da série histórica.

Em relação aos dados reais, foi apurado que o conjunto de ativos atinge o alvo estabelecido, em média, em 40,99% dos pregões. Já a classificação da SVM conseguiu um resultado médio de 48,37%, ou seja, uma melhora de 17,99% entre a decisão trivial e o classificador SVM proposto. Verificou-se, por meio do teste de Wilcoxon, que as diferentes performances são estatisticamente diferentes para um nível de significância de 95%.

A Figura 5 apresenta a comparação entre os dados históricos e a SVM, para o conjunto de 50 ativos da amostra.



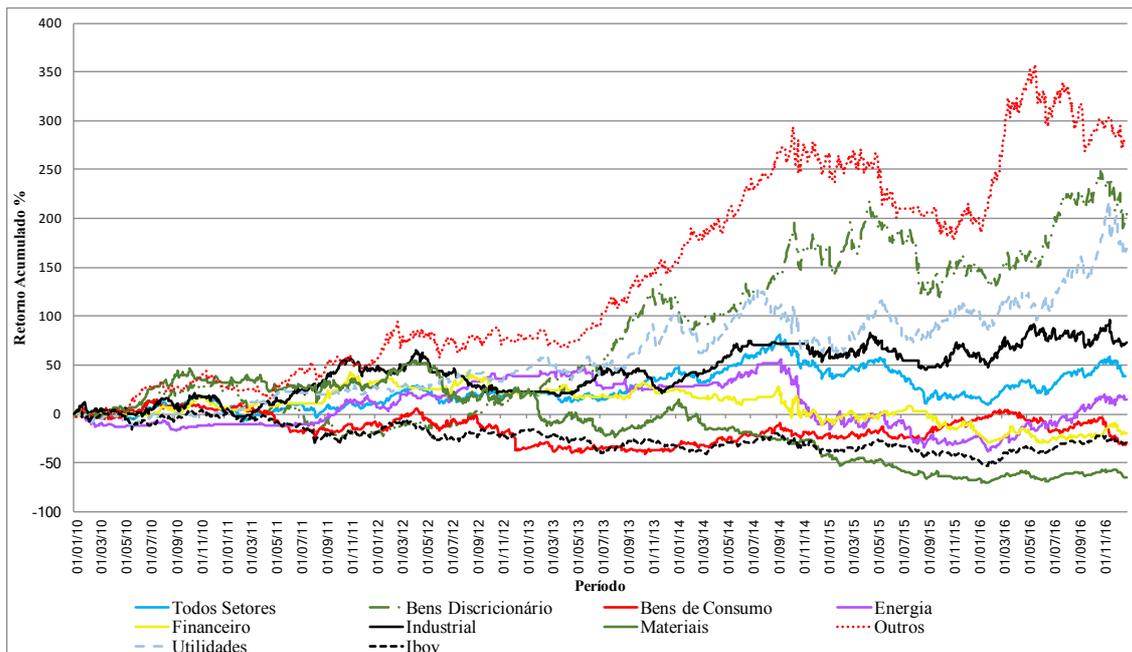
**Figura 5 - Performance do classificador SVM**

Na sequência comparou-se a performance preditiva da SVM segmentando os ativos por grupos de atividade econômica. Pode-se verificar por meio da Tabela 3 que, a SVM alcançou uma melhor performance que os dados históricos. Quanto as melhores classificações, destacamos os setores de Energia, Materiais e Utilidade Pública. Não se verifica uma discrepância muito alta de classificação entre os ativos pertencentes aos seus respectivos grupos econômicos. A Tabela 4 disposta no Apêndice tem maiores detalhes quanto a retorno por ativos e seu desempenho por carteira setorial.

**Tabela 3- Performance por setor econômico**

Setor	Série Histórica			Classificador SVM			Variação %
	Média %	Desvio-Padrão	CV	Média %	Desvio-Padrão	CV	
Todos Setores	40,99	6,01	0,15	48,37	6,78	0,14	17,99
Discricionário	46,77	3,96	0,08	51,07	4,01	0,08	9,21
Consumo	40,40	8,20	0,20	45,33	7,45	0,16	12,20
Energia	38,71	4,54	0,12	49,69	5,42	0,11	28,38
Financeiro	35,66	3,14	0,09	40,53	4,78	0,12	13,68
Industrial	40,84	3,40	0,08	46,09	4,30	0,09	12,85
Materiais	43,47	5,70	0,13	55,75	4,13	0,07	28,26
Outros	41,18	5,39	0,13	45,28	4,83	0,11	9,96
Util Public.	39,48	3,82	0,10	49,36	4,05	0,08	25,03

A fase seguinte da pesquisa focou na análise de rentabilidade das decisões de investimento baseadas na técnica SVM para classificação do movimento de alta do preço das ações. Assim, conforme destacado pela Figura 6, o retorno acumulado das operações de investimento, ao longo do período de análise, sem distinção de setor econômico apresentou uma rentabilidade bem superior ao *baseline* Ibov, 38,02% contra -27,15%. Por sua vez, o retorno acumulado para o setor Outros foi de 278,55%, na ponta oposta figurou o setor Materiais com retorno negativo de 65,43%%. Os resultados finais dos demais foram: Bens de Consumo Discricionários (206,99%); Bens de Consumo (-29,23%); Energia (14,34%); Financeiro (-18,48%); Industrial (73,81%); Utilidade Pública (169,03%).



**Figura 6 – Retorno acumulado da carteira por setor**

Se na Figura 6 foi apresentado os dados de performance diária, na Tabela 5 o resultado é analisado numa perspectiva mensal. Os ativos dos setores de Bens de Consumo Discricionários e Outros apresentaram retornos médios mensais bem superiores aos demais, além dos menores coeficientes de variação de Pearson, o que indica uma menor volatilidade.

**Tabela 5- Retorno mensal por setor econômico**

Setor	Média %	Desvio-Padrão	CV	Máximo	Mínimo
Todos Setores	0,53	5,34	10,11	10,86	-12,37
Discricionário	1,63	7,60	4,65	19,99	-20,91
Consumo	-0,23	5,87	-25,23	11,37	-19,04
Energia	0,41	6,90	16,94	21,61	-24,13
Financeiro	-0,09	5,52	-62,55	11,79	-13,05
Industrial	0,80	5,29	6,62	13,79	-11,65
Materiais	-0,97	7,43	-7,66	15,67	-21,42
Outros	1,76	5,75	3,27	18,12	-13,40
Util Public.	0,53	5,34	10,11	10,86	-12,37

Quanto a cardinalidade das carteiras de investimento, que implica diretamente nos custos operacionais, percebe-se por meio da Tabela 6 que a carteira composta por ativos de todos setores tem cardinalidade média de 14,37 ativos, ou seja, a carteira é composta em média por 28,74% dos 50 ativos disponíveis na amostra. A cardinalidade média do setor Outros, melhor índice rentabilidade entre os setores, foi de 1,65 ativos, não foi a carteira de menor cardinalidade, mas foi bem menor que a carteira formada por ativos de todos os setores. Esse é um bom indicador, já que a cardinalidade da carteira impacta diretamente no custo operacional da mesma. Sendo assim, o resultado apresentado corrobora ainda mais para um melhor resultado líquido.

**Tabela 6- Cardinalidade das carteiras**

Carteira	Média	Desvio-Padrão	CV	Máximo	Mínimo
Todos Setores	14,37	9,19	0,64	40	1
Discricionário	2,49	1,46	0,59	6	0
Consumo	2,00	1,40	0,70	7	0
Energia	1,03	1,14	1,11	4	0
Financeiro	1,00	1,36	1,37	6	0
Industrial	1,01	1,05	1,04	4	0
Materiais	3,08	2,72	0,88	9	0
Outros	1,65	1,38	0,84	6	0
Util. Pública	2,11	2,08	0,99	8	0

Na Tabela 7 percebe-se que dos 1.733 dias de negociações analisados, as estratégias baseadas nos ativos dos setores Financeiro e Energia foram as que menos apresentaram *trading* diário. A que se destacar os setores Bens de Consumo Discricionários, Serviços de Utilidade Pública e Outros, que tiveram presentes em mais de 70% dos dias de negociação, mas que alcançaram um ganho médio bem superior aos demais. Destaque também para o menor coeficiente de variação de Pearson da carteira baseada no setor Outros, mostrando uma maior confiança do resultado médio se comparado aos outros setores econômicos.

**Tabela 7- Tradings por setor**

Carteira	Dias Neg.	Dias Neg. %	Neg. Compra	Res. Méd %	Desvio-Padrão	CV	Máximo %	Mínimo %
Todos Setores	1.733	100,00	24.911	0,02	1,11	44,68	4,06	-7,76
Discricionário	1.601	92,38	4.322	0,08	1,62	19,44	4,70	-8,60
Consumo	1.485	85,69	3.473	-0,01	1,67	182,15	3,21	-16,19
Energia	935	53,95	1.780	0,03	1,95	57,56	4,02	-11,71
Financeiro	860	49,62	1.727	-0,01	1,93	380,83	18,20	-13,43
Industrial	1.058	61,05	1.753	0,07	1,58	24,26	2,51	-6,77
Materiais	1.402	80,90	5.332	-0,06	1,80	30,49	4,82	-10,05
Outros	1.313	75,76	2.862	0,11	1,52	13,43	3,32	-9,01
Util. Pública	1.266	73,05	3.662	0,09	1,58	17,34	14,73	-8,22

Na Tabela 8 para uma melhor análise, o resultado das operações que geraram lucro e prejuízo. Para todos os tipos de carteiras verifica-se que o valor do ganho médio das operações foi inferior ao valor da perda média. Especificamente, o setor Outros apesar do valor da perda média ser superior ao valor do ganho médio, o número de operações de operações de ganho compensou esse déficit. Ou seja, para cada 1 negociação de perda é preciso ter 1,35 negociações com ganho para equilibrar o resultado, só há geração de lucro caso a proporção seja de negociações de ganho superem as de prejuízo em 1,35, como essa relação foi 18,86 o resultado final favorável ao investidor.

**Tabela 8- Resultado dos tradings por setor**

Setor	% Trading > 0	Ganho Médio (%)	% Trading < 0	Perda Média (%)	Ratio Loss/Gain	Ratio G/L Trading
Todos Setores	88,31	0,75	11,69	0,97	1,30	7,56
Discricionário	74,91	1,10	25,09	1,56	1,41	2,98
Consumo	37,26	1,11	62,74	1,50	1,35	0,59
Energia	40,13	1,26	59,87	1,92	1,52	0,67
Financeiro	49,21	1,24	50,79	1,74	1,40	0,97
Industrial	94,14	1,15	5,86	1,52	1,32	16,05
Materiais	46,48	1,14	53,52	1,83	1,62	0,87
Outros	94,96	1,09	5,04	1,48	1,35	18,86
Util Public.	79,63	1,07	20,37	1,48	1,38	3,91

Com relação a exposição de risco, a medida de máximo *dropdown* corrobora para avaliar as estratégias de investimento. Assim, percebe-se a partir da Tabela 9 que, investimento

baseado no setor Outros foi o com menor exposição ao risco um *dropdown* máximo de 21,00%, bem como o menor *dropdown* médio.

**Tabela 9- Exposição ao risco por setor**

Carteira	Máx Dropdown (%)	Dropdown Médio (%)	Desvio-Padrão	CV
Todos Setores	37,18	14,21	11,68	0,82
Discrecionário	40,22	11,01	11,85	1,08
Consumo	44,70	17,94	18,99	1,06
Energia	58,14	14,33	18,58	1,30
Financeiro	48,32	12,71	16,43	1,29
Industrial	23,61	10,70	7,15	0,67
Materiais	79,54	30,72	34,73	1,13
Outros	21,00	6,94	5,99	0,86
Util. Pública	37,18	12,08	11,96	0,99

## 5 CONSIDERAÇÕES FINAIS

A instigante dúvida acerca da capacidade de se prever retornos futuros de ativos financeiros parece perdurar na pauta dos acadêmicos. Os estudos indicam que aqueles conduzidos pela aplicação de técnicas de *machine learnig* parecem ser bastante promissores. Seguindo essa linha de pensamento, definimos explorar no presente trabalho a técnica de *support vector machine*, com o intuito de criar um classificador de movimentos futuros dos preços dos ativos que indicasse uma possível situação de lucro.

O resultado da pesquisa apresentou um resultado bastante interessante à medida que o retorno de investimento baseado na SVM foi bem superior ao *baseline* (IBovespa). Aponta-se também que, construir um classificador com ativos homogêneos quanto ao setor econômico de atuação, proporcionou um melhor índice de rentabilidade, menor cardinalidade média das carteiras de investimento, menor máximo *dropdown*. Ou seja, a composição do classificador para decisões de investimento a partir dos ativos do setor Outros (composto por empresas dos setores de Imóveis, Tecnologia da Informação, Atendimento a Saúde e Telecomunicações) resultou numa rentabilidade 632,64% maior, numa cardinalidade média 88.51% menor e um *dropdown* máximo 43,52% menor, se comparar ao classificador que não faz distinção de setor econômico. Resumindo o classificador SVM com base nos ativos do setor Outros resultou numa maior rentabilidade, menor custo operacional e menor exposição ao risco.

Desta forma, um modelo de decisão de investimento baseado em técnicas de inteligência computacionais denota um caráter promissor, ensejando grandes possibilidades de continuidade dos estudos. Deixamos, como proposta de estudos futuros, a sugestão de inclusão de custos de transação, implementação de *stop-loss* para as operações e testes com outras técnicas de inteligência computacional para previsão do movimento dos retornos dos ativos.

## REFERÊNCIAS

- Awad, M., & Khanna, R. (2015). *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Berkeley, CA: Apress.
- Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42, 7046–7056.

- Brabazon, A., & O'Neill, M. (2006). *Biologically Inspired Algorithms for Financial Modelling*. Berlin: Springer.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2, 121–167.
- Ding, Y., Song, X., & Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications*, 34, 3081–3089.
- Fenghua, W., Jihong, X., Zhifang, H., & Xu, G. (2014). Stock Price Prediction based on SSA and SVM. *Procedia Computer Science*, 31, 625–631.
- Huang, C., Yang, D., & Chuang, Y. (2008). Application of wrapper approach and composite classifier to the stock trend prediction. *Expert Systems with Applications*, 34, 2870–2878.
- Huang, W., Nakamoria, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research*, 32, 2513–2522.
- Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38, 5311–5319.
- Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55, 307–319.
- Lee, M. C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, 36, 10896–10904.
- Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., ... Deng, X. (2016). Empirical analysis: stock market prediction via extreme learning machine. *Neural Computing and Applications*, 27, 67–78.
- Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. *The Journal of Finance*, 55, 1705–1765.
- Luo, L., & Chen, X. (2013). Integrating piecewise linear representation and weighted support vector machine for stock trading signal prediction. *Applied Soft Computing*, 13, 806–816.
- Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17, 59–82.
- Mandelbrot, B., & Hudson, R. L. (2004). *The (Mis)behavior of Markets: A Fractal View of Financial Turbulence*. The Mathematical Intelligencer. New York: Basic Books.
- Ni, L.-P., Ni, Z.-W., & Gao, Y.-Z. (2011). Stock trend prediction based on fractal feature selection and support vector machine. *Expert Systems with Applications*, 38, 5569–5576.
- Pan, Y., Xiao, Z., Wang, X., & Yang, D. (2017). A multiple support vector machine approach to stock index forecasting with mixed frequency sampling. *Knowledge-Based Systems*, 122, 90–102.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*, 42, 259–268.
- Sands, T. M., Tayal, D., Morris, M. E., & Monteiro, S. T. (2015). Robust Stock Value Prediction using Support Vector Machines with Particle Swarm Optimization. *2015 IEEE Congress on Evolutionary Computation (CEC)*, 3327–3331.

- Sheta, A. F., Ahmed, S. E. M., & Faris, H. (2015). A Comparison between Regression, Artificial Neural Networks and SVM for Predicting Stock Market Index. *International Journal of Advanced Research in Artificial Intelligence*, 4, 55–63.
- Tan, T. Z., Quek, C., & Ng, G. S. (2007). Biological brain-inspired genetic complementary learning for stock market and bank failure prediction. *Computational Intelligence*, 23, 236–261.
- Thenmozhi, M., & Sarath Chand, G. (2016). Forecasting stock returns based on information transmission across global markets using support vector machines. *Neural Computing and Applications*, 27, 805–824.
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10, 988–999.
- Wen, Q., Yang, Z., Song, Y., & Jia, P. (2010). Automatic stock decision support system based on box theory and SVM algorithm. *Expert Systems with Applications*, 37, 1015–1022.
- Yu, H., Chen, R., & Zhang, G. (2014). A SVM Stock Selection Model within PCA. *Procedia Computer Science*, 31, 406–412.
- Zbikowski, K. (2015). Using Volume Weighted Support Vector Machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. *Expert Systems with Applications*, 42, 1797–1805.
- Zhiqiang, G., Huaiqing, W., & Quan, L. (2013). Financial time series forecasting using LPP and SVM optimized by PSO. *Soft Computing*, 17, 805–818.

## APÊNDICE

**Tabela 2- Amostra da pesquisa**

Ticker	Empresa	Setor
ABEV3	AMBEV SA	Produtos básicos de consumo
BBAS3	BANCO DO BRASIL S.A.	Serviços financeiros
BBDC3	BANCO BRADESCO S.A.	Serviços financeiros
BBDC4	BANCO BRADESCO SA-PREF	Serviços financeiros
BRAP4	BRADESPAR SA -PREF	Materiais
BRFS3	BRF SA	Produtos básicos de consumo
BRKM5	BRASKEM SA-PREF A	Materiais
BRML3	BR MALLS PARTICIPACOES SA	Outros
BVMF3	BM&FBOVESPA SA	Serviços financeiros
CCRO3	CCR SA	Industriais
CIEL3	CIELO SA	Outros
CMIG4	CIA ENERGETICA MINAS GER-PRF	Serviços de utilidade pública
CPFE3	CPFL ENERGIA SA	Serviços de utilidade pública
CPL6	CIA PARANAENSE DE ENERGI-PFB	Serviços de utilidade pública
CSAN3	COSAN SA INDUSTRIA COMERCIO	Energia
CSNA3	CIA SIDERURGICA NACIONAL SA	Materiais
CYRE3	CYRELA BRAZIL REALTY SA EMP	Bens de consumo discricionário
EGIE3	ENGIE BRASIL ENERGIA SA	Serviços de utilidade pública
ELET3	CENTRAIS ELETRICAS BRASILIER	Serviços de utilidade pública
EMBR3	EMBRAER SA	Industriais
ENBR3	EDP - ENERGIAS DO BRASIL SA	Serviços de utilidade pública
EQTL3	EQUATORIAL ENERGIA SA - ORD	Serviços de utilidade pública
ESTC3	ESTACIO PARTICIPACOES SA	Bens de consumo discricionário
GGBR4	GERDAU SA-PREF	Materiais
GOAU4	METALURGICA GERDAU SA-PREF	Materiais
HYPE3	HYPERMARCAS SA	Outros
ITSA4	ITAUSA-INVESTIMENTOS ITAU-PR	Serviços financeiros
ITUB4	ITAU UNIBANCO HOLDING S-PREF	Serviços financeiros
JBSS3	JBS SA	Produtos básicos de consumo
KROT3	KROTON EDUCACIONAL SA	Bens de consumo discricionário
LAME4	LOJAS AMERICANAS SA-PREF	Bens de consumo discricionário
LREN3	LOJAS RENNER S.A.	Bens de consumo discricionário
MRFG3	MARFRIG GLOBAL FOODS SA	Produtos básicos de consumo
MRVE3	MRV ENGENHARIA	Bens de consumo discricionário
MULT3	MULTIPLAN EMPREENDIMENTOS	Outros
NATU3	NATURA COSMETICOS SA	Produtos básicos de consumo
PCAR4	CIA BRASILEIRA DE DIS-PREF	Produtos básicos de consumo
PETR3	PETROBRAS - PETROLEO BRAS	Energia
PETR4	PETROBRAS - PETROLEO BRAS-PR	Energia
RADL3	RAIA DROGASIL SA	Produtos básicos de consumo
RENT3	LOCALIZA RENT A CAR	Industriais
SBSP3	CIA SANEAMENTO BASICO DE SP	Serviços de utilidade pública
SUZB5	SUZANO PAPEL E CELULO-PREF A	Materiais
TIMP3	TIM PARTICIPACOES SA	Outros
UGPA3	ULTRAPAR PARTICIPACOES SA	Energia
USIM5	USINAS SIDER MINAS GER-PF A	Materiais
VALE3	VALE SA	Materiais
VALE5	VALE SA-PREF	Materiais
VIVT4	TELEFONICA BRASIL S.A.-PREF	Outros
WEGE3	WEG SA	Industriais

Nota: Todos os setores que não tinham mais de 2 ativos foram agrupados em Outros. Para conhecimento destaca-se aqui os setores de origem de cada empresa reclassificada. BRML3 (Imóveis); CIEL3 (Tecnologia da Informação); HYPE3 (Atendimento à Saúde); MULT3 (Imóveis); TIMP3 (Serviços de Telecomunicações); VIVT4 (Serviços de Telecomunicações).

**Tabela 4 – Retorno por ativo destacado por carteiras setoriais**

Ticker	Todos Setores	Bens Discricionári	Bens de Consumo	Energia	Financeiro	Industrial	Materiais	Outros	Utilidades
ABEV3	0,82	0,00	2,19	0,00	0,00	0,00	0,00	0,00	0,00
BBAS3	-1,30	0,00	0,00	0,00	-25,42	0,00	0,00	0,00	0,00
BBDC3	1,20	0,00	0,00	0,00	8,67	0,00	0,00	0,00	0,00
BBDC4	-0,11	0,00	0,00	0,00	6,97	0,00	0,00	0,00	0,00
BRAP4	-1,52	0,00	0,00	0,00	0,00	0,00	-7,88	0,00	0,00
BRFS3	6,73	0,00	17,08	0,00	0,00	0,00	0,00	0,00	0,00
BRKM5	4,99	0,00	0,00	0,00	0,00	0,00	21,31	0,00	0,00
BRML3	0,63	0,00	0,00	0,00	0,00	0,00	0,00	1,25	0,00
BVMF3	1,92	0,00	0,00	0,00	1,12	0,00	0,00	0,00	0,00
CCRO3	3,49	0,00	0,00	0,00	0,00	29,44	0,00	0,00	0,00
CIEL3	1,54	0,00	0,00	0,00	0,00	0,00	0,00	-6,38	0,00
CMIG4	-0,37	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-6,34
CPFE3	3,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	18,04
CPLE6	2,87	0,00	0,00	0,00	0,00	0,00	0,00	0,00	19,57
CSAN3	3,19	0,00	0,00	42,76	0,00	0,00	0,00	0,00	0,00
CSNA3	-6,27	0,00	0,00	0,00	0,00	0,00	-16,31	0,00	0,00
CYRE3	-0,15	21,47	0,00	0,00	0,00	0,00	0,00	0,00	0,00
EGIE3	4,56	0,00	0,00	0,00	0,00	0,00	0,00	0,00	19,05
ELET3	-4,39	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-7,67
EMBR3	-2,31	0,00	0,00	0,00	0,00	-37,20	0,00	0,00	0,00
ENBR3	1,58	0,00	0,00	0,00	0,00	0,00	0,00	0,00	11,51
EQTL3	4,82	0,00	0,00	0,00	0,00	0,00	0,00	0,00	21,23
ESTC3	-0,47	-12,17	0,00	0,00	0,00	0,00	0,00	0,00	0,00
GGBR4	-1,28	0,00	0,00	0,00	0,00	0,00	-2,53	0,00	0,00
GOAU4	-1,06	0,00	0,00	0,00	0,00	0,00	-4,77	0,00	0,00
HYPE3	11,12	0,00	0,00	0,00	0,00	0,00	0,00	84,80	0,00
ITSA4	0,06	0,00	0,00	0,00	-2,05	0,00	0,00	0,00	0,00
ITUB4	-0,28	0,00	0,00	0,00	6,37	0,00	0,00	0,00	0,00
JBSS3	-1,42	0,00	-13,70	0,00	0,00	0,00	0,00	0,00	0,00
KROT3	3,77	29,29	0,00	0,00	0,00	0,00	0,00	0,00	0,00
LAME4	5,48	20,84	0,00	0,00	0,00	0,00	0,00	0,00	0,00
LREN3	4,89	24,72	0,00	0,00	0,00	0,00	0,00	0,00	0,00
MRFG3	-3,71	0,00	-14,53	0,00	0,00	0,00	0,00	0,00	0,00
MRVE3	-7,54	49,47	0,00	0,00	0,00	0,00	0,00	0,00	0,00
MULT3	9,86	0,00	0,00	0,00	0,00	0,00	0,00	57,17	0,00
NATU3	0,97	0,00	-3,12	0,00	0,00	0,00	0,00	0,00	0,00
PCAR4	-2,46	0,00	-23,02	0,00	0,00	0,00	0,00	0,00	0,00
PETR3	0,78	0,00	0,00	11,94	0,00	0,00	0,00	0,00	0,00
PETR4	-6,57	0,00	0,00	-46,28	0,00	0,00	0,00	0,00	0,00
RADL3	2,56	0,00	21,56	0,00	0,00	0,00	0,00	0,00	0,00
RENT3	3,13	0,00	0,00	0,00	0,00	12,09	0,00	0,00	0,00
SBSP3	4,93	0,00	0,00	0,00	0,00	0,00	0,00	0,00	39,42
SUZB5	-0,10	0,00	0,00	0,00	0,00	0,00	7,67	0,00	0,00
TIMP3	-0,20	0,00	0,00	0,00	0,00	0,00	0,00	-1,74	0,00
UGPA3	2,50	0,00	0,00	23,19	0,00	0,00	0,00	0,00	0,00
USIM5	-12,20	0,00	0,00	0,00	0,00	0,00	-57,40	0,00	0,00
VALE3	-2,25	0,00	0,00	0,00	0,00	0,00	-9,33	0,00	0,00
VALE5	-4,39	0,00	0,00	0,00	0,00	0,00	-13,72	0,00	0,00
VIVT4	2,77	0,00	0,00	0,00	0,00	0,00	0,00	13,49	0,00
WEGE3	8,76	0,00	0,00	0,00	0,00	64,24	0,00	0,00	0,00